Online Appendix

Methods, Robustness, Sensitivity, and Reproducibility

Supplement to *Vaccines Don't Cause Autism*

Jordan Lasker - Texas Tech University - 28 October 2026

These analyses were not preregistered. Analysis choices were documented during the project, including after earlier results had been examined. Rule files and hash records document computational settings and provenance; they do not establish preregistration.

This appendix accompanies the paper and preserves the source-specific methods, positive and negative controls, failed diagnostics, effect-injection experiments, uncertainty assumptions and reproducibility records. Numerical estimates remain on the scale of their own exposure contrast and outcome system.

Sections S1-S73 contain the established data, design, outcome-specific and synthesis material. Sections S74-S79 provide aggregate event-study methods and reference analyses, cancer models, additional design-specific evidence, matched contrasts, receipt-conversion checks and maximal-inclusion results. Historical analyses retain their stated editions and are not independent confirmations.

EA1 contains the full effect-recovery plot moved from the main paper. EA2 records analysis-edition and publication provenance. EA3 contains the full matched-control event-time display. EA4 provides detailed result tables, recovery and placebo tests, numerical verification and unresolved source or historical discrepancies. EA5 contains additional model comparisons and source-specific sensitivities. EA6 defines how estimates, intervals, controls, interactions, timing and diagnostic checks are interpreted throughout this appendix, including the retained reference analyses. The main paper reports the original public A7 analysis snapshot: 39,753 comparisons around 404 changes in 40 countries.

The PDF bookmarks provide direct access to sections, tables and figures. The original source catalog remains included. The later private timing-qualified candidate is not substituted into this edition.

S1. Purpose and reporting protocol

Interpretation of PCV7 calculations is governed by the exposure-aligned audit in Section S52. The spline's model-native intervals and numerical calibration exercises do not supply individual vaccine-risk bounds. The common-direction synthesis containing that coefficient likewise does not identify a probability of generalized vaccine harm. Raw uptake changes, numerical reproduction and prior sensitivity do not establish the required exposure-to-estimator mapping.

This appendix reports robustness, sensitivity, and source documentation for *Vaccines Don't Cause Autism*. Robustness is an identification requirement, not a search for favorable specifications. The analysis therefore reports null, positive, inverse, unstable, and failed results; presents the implemented estimator and trend variants together; and distinguishes a diagnostic that passes from a design that is causally identified. A passing placebo cannot repair weak exposure measurement, a short preperiod, a contemporaneous outcome-system change, or an ecological estimand. Source-level integration coverage is documented in S55-S56 and S73. Section S67 defines the full 17-context synthesis, denominator-separated local associations and native-scale published-study summaries. The restricted 15-context models, their calibration and power calculations in S59-S66 retain their stated samples and assumptions; their numerical results are not the expanded hierarchy's operating characteristics.

The analysis has ten layers: exposure-positive controls; claim-relevant comparator outcomes; fake-event, future-exposure, and assignment placebos; alternative estimators; flexible trend and age-period-cohort specifications; influence, permutation, and small-cluster inference; model-rank and leverage diagnostics; preperiod-only complexity selection; familywise event-study bands; and design-conditional power, equivalence, and Bayes-factor calibration. Machine-readable files 128 through 143 contain the results, estimand translations, and eligibility classifications.

S2. Exposure-positive controls

The first question is whether the event and cohort definitions detect changes in vaccine exposure where they should. Eleven of twelve exposure or internal-measurement checks move in the expected direction. The U.S. NIS panel records a 34.8-percentage-point PCV7 increase between proxy birth cohorts 1999 and 2000 and an 18.4-point rotavirus increase between cohorts 2005 and 2006. Published German first stages range from 43.8 to 72.9 points; Italy's largest adjacent-cohort changes are 16.3 points for varicella and 36.7 points for rotavirus; Swedish constructed programme shares jump by 55.2 points for PCV, 97.3 for HepB, and 84.0 for rotavirus. These checks validate exposure timing and measurement, not effects on autism.

The one limited internal positive control is SchoolVaxView: a one-point increase in kindergarten exemptions predicts 1.16 points lower MMR coverage after state and cohort fixed effects, as expected, but the cluster-robust p-value is 0.129. We retain it as directionally coherent but imprecise rather than relabeling it a pass.

A final same-design positive-control search uses two official CDC Active Bacterial Core surveillance releases. The site-by-age-by-year serotype file contains 30,049 rows across ten surveillance sites from 1998 through 2024, and the national age-by-year file contains 1,677 rows from 1997 through 2024. Neither provides birth cohort, year-specific site denominators, or a stable site-catchment crosswalk; CDC documents changing catchments and surveillance methods. Five required gates therefore fail, and file 2740 records `FAIL_CLOSED_NO_SAME_DESIGN_ESTIMATOR` instead of manufacturing a state-by-cohort regression. The national timing calibration nevertheless shows the expected post-PCV7 age pattern: comparing 1998-2000 with 2001-2003, invasive pneumococcal disease rates decline 72.2 percent below age one, 76.2 percent at age one, 43.7 percent at ages two to four, and 18.0 percent at ages five to seventeen. This is partial exposure-coherence evidence, not a validation of the exact PCV7 intensity instrument and not evidence about autism.

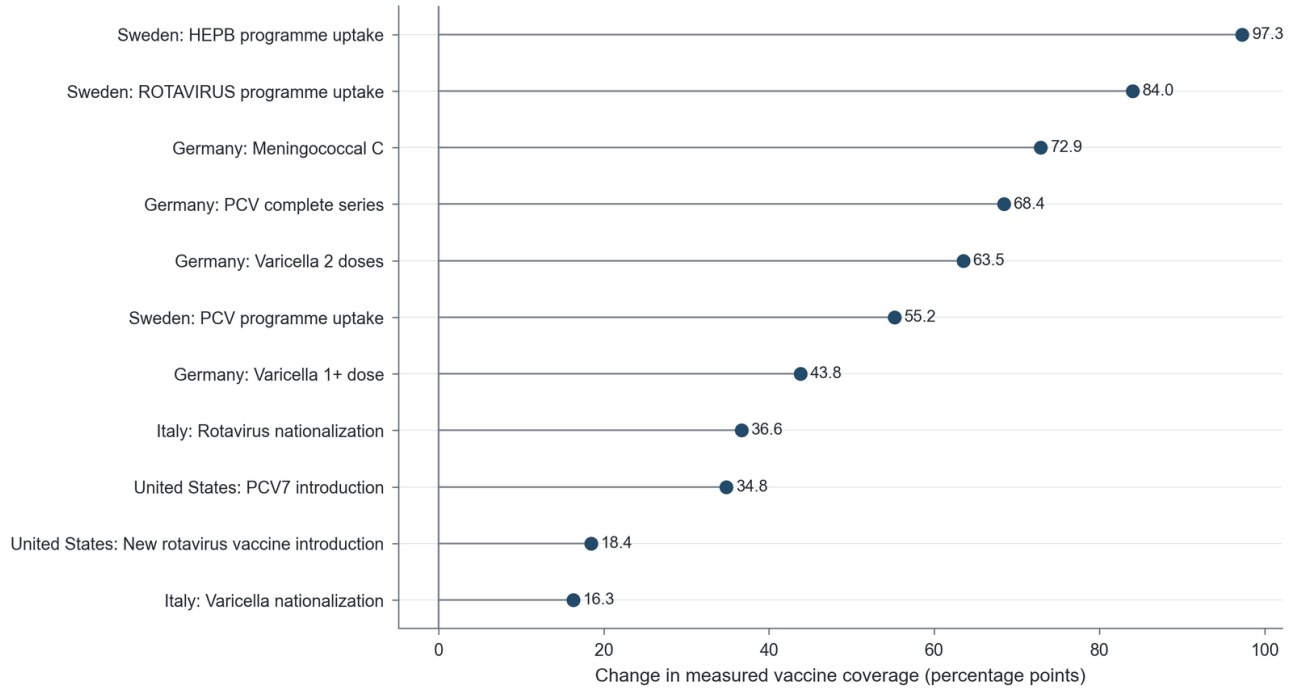


Figure S1. Exposure-positive controls and measured first stages

Table S1. Exposure-positive controls

Setting	Exposure/check	Before	After	Change	p	Assessment
United States	PCV7 introduction	34.4	69.2	34.8 pp	<0.001	Pass
United States	New rotavirus vaccine introduction	0.0	18.5	18.4 pp	<0.001	Pass
U.S. SchoolVax View	Exemption-to-MMR measurement check	-	-	-1.2 pp	0.129	Expected sign; imprecise
Germany	PCV complete series	3.9	72.3	68.4 pp	-	Pass
Germany	Meningococcal C	8.2	81.1	72.9 pp	-	Pass
Germany	Varicella 1+ dose	43.2	87.0	43.8 pp	-	Pass
Germany	Varicella 2 doses	0.6	64.1	63.5 pp	-	Pass
Italy	Varicella nationalization	74.2	90.5	16.3 pp	-	Pass
Italy	Rotavirus nationalization	26.1	62.8	36.6 pp	-	Pass
Sweden	PCV programme uptake	0.0	55.2	55.2 pp	-	Pass
Sweden	HEPB programme uptake	0.0	97.3	97.3 pp	-	Pass
Sweden	ROTAVIRUS programme uptake	0.0	84.0	84.0 pp	-	Pass

Notes: Positive controls validate exposure timing or internal measurement coherence. They do not identify effects on autism.

S3. Alternative estimators and functional forms

Thirty-one estimator/specification fits compare Poisson pseudo-maximum likelihood, population-weighted and unweighted log-rate regressions, independent and exchangeable Poisson GEE, differential linear and quadratic trends, endpoint trimming, flexible cohort splines, and outcome-system-break adjustments. No fit failed computationally.

For U.S. IDEA PCV7, point estimates range from RR 0.988 to 1.023 per one-standard-deviation higher early uptake, and every confidence interval includes one. SchoolVaxView MMR estimates range from RR 0.959 to 0.989 per ten percentage points, again with every interval including one. Its state-specific linear-trend fit is a sensitivity analysis rather than an identification improvement: with three outcome cohorts, the model matrix has rank 101 in 142 observations. ADDM PCV7 estimates range from RR 0.947 to 1.072 per six-site SD of early uptake (2.35 points; 0.793 to 1.345 per ten points) and move across one with only six sites, so this panel is labeled a small-cluster sensitivity rather than an independent replication. Scotland ranges from RR 0.844 to 1.020 with broad intervals. Wales ranges from RR 0.771 to 1.051 in the alternative-estimator audit; the separate one-step-up model reaches RR 1.692 (1.033 to 2.771), but a line through its two preperiods is exact interpolation and has no independent pretrend-validation content.

The PCV7 state ranking is correlated with other vaccine-uptake rankings, so a final joint product-attribution model enters PCV7, MMR, and varicella simultaneously while retaining the preferred PCV intensity-specific cohort path. PCV7 is RR 1.020 (0.950-1.095), MMR is 0.997 (0.959-1.037), and varicella is 0.993 (0.940-1.049); the conditional PCV MDE80 is 1.105. About 0.75 standard deviations of PCV intensity remain after residualizing it on MMR and varicella, corresponding to variance-inflation factor 1.79. This exercise separates correlated state coverage rankings only under the maintained cohort path. It does not create product-specific random assignment (`2882_flagship_product_attribution_horserace.csv`; `2883_flagship_product_attribution_horserace_validation.csv`).

Holding the flexible pooled-age English specification fixed, Poisson, WLS, OLS, and GEE estimates are similarly inverse (RR 0.937 to 0.950). This is estimator robustness, not identification robustness: the age-eight specification is near one, and alternative cohort/APC restrictions materially change the result. The supplement therefore classifies England as high-power but trend-form fragile.

A final held-input England audit expands this check without promoting the national step to a causal design. Across every contiguous cutoff with at least three cohorts on each side, the 2014 policy cutoff gives RR 0.961 (0.928-0.996), but its plus-one rank among thirteen eligible cutoffs is 0.214. That rank is descriptive, not a randomization p-value, because national cohort cutoffs are ordered and nonexchangeable. Thirty-six leave-one-cohort, -age, or -year fits, eight Poisson/log-rate/negative-binomial functional forms, cohort-score Newey-West lags zero through three, the full legal nearest-five count interval, and 250 fixed-seed count-jitter draws leave the inverse location broadly intact. These are influence, functional-form, serial-dependence, and measurement-perturbation diagnostics; the absence of an untreated geography and sensitivity to the cohort path keep causal weight at zero (files 2880-2889).

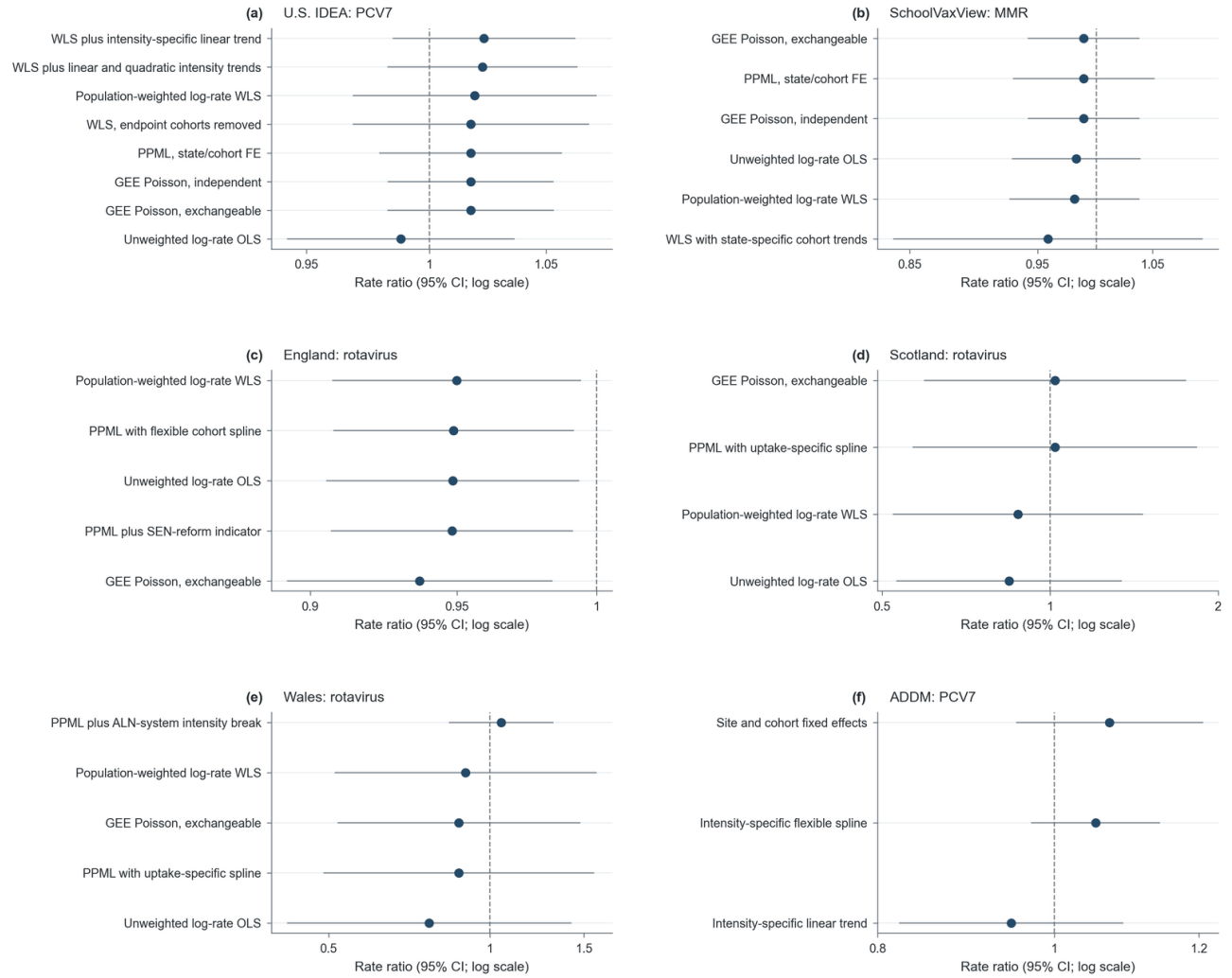


Figure S2. Alternative estimators and functional-form specifications

Table S2. Complete alternative-estimator and specification results

U.S. IDEA PCV7

Specification	Estimator	RR	95% CI	p	N	Clusters
PPML, state/cohort FE	ppml_cluster	1.017	0.979-1.057	0.368	353	51
Population-weighted log-rate WLS	wls_log_rate_cluster	1.019	0.969-1.072	0.458	353	51
Unweighted log-rate OLS	ols_log_rate_cluster	0.988	0.942-1.036	0.617	353	51
GEE Poisson, independent	gee_poisson_independent	1.017	0.983-1.053	0.321	353	51
GEE Poisson, exchangeable	gee_poisson_exchangeable	1.017	0.983-1.053	0.323	353	51
WLS plus intensity-specific linear trend	wls_log_rate_cluster	1.023	0.985-1.063	0.234	353	51
WLS plus linear and quadratic intensity trends	wls_log_rate_cluster	1.022	0.983-1.064	0.268	353	51
WLS, endpoint cohorts removed	wls_log_rate_cluster	1.018	0.968-1.069	0.483	251	51

SchoolVaxView MMR

Specification	Estimator	RR	95% CI	p	N	Clusters
PPML, state/cohort FE	ppml_cluster	0.989	0.930-1.052	0.718	142	50
Population-weighted log-rate WLS	wls_log_rate_cluster	0.981	0.927-1.038	0.498	142	50
Unweighted log-rate OLS	ols_log_rate_cluster	0.983	0.929-1.039	0.532	142	50
GEE Poisson, independent	gee_poisson_independent	0.989	0.942-1.038	0.646	142	50
GEE Poisson, exchangeable	gee_poisson_exchangeable	0.989	0.942-1.038	0.646	142	50
WLS with state-specific cohort trends	wls_log_rate_cluster	0.959	0.838-1.097	0.531	142	50

England rotavirus

Specification	Estimator	RR	95% CI	p	N	Clusters
PPML with flexible cohort spline	ppml_cluster	0.949	0.908-0.992	0.022	128	23
Population-weighted log-rate WLS	wls_log_rate_cluster	0.950	0.907-0.994	0.029	128	23
Unweighted log-rate OLS	ols_log_rate_cluster	0.948	0.905-0.994	0.028	128	23
GEE Poisson, exchangeable	gee_poisson_exchangeable	0.937	0.892-0.984	0.011	128	23
PPML plus SEN-reform indicator	ppml_cluster	0.948	0.907-0.991	0.021	128	23

Scotland rotavirus

Specification	Estimator	RR	95% CI	p	N	Clusters
PPML with uptake-specific spline	ppml_cluster	1.020	0.567-1.833	0.946	512	32
Population-weighted log-rate WLS	wls_log_rate_cluster	0.876	0.523-1.467	0.603	512	32
Unweighted log-rate OLS	ols_log_rate_cluster	0.844	0.530-1.345	0.464	512	32
GEE Poisson, exchangeable	gee_poisson_exchangeable	1.020	0.594-1.753	0.940	512	32

Wales rotavirus

Specification	Estimator	RR	95% CI	p	N	Clusters
PPML with uptake-specific spline	ppml_cluster	0.875	0.489-1.566	0.639	220	22
Population-weighted log-rate WLS	wls_log_rate_cluster	0.901	0.513-1.583	0.705	220	22
Unweighted log-rate OLS	ols_log_rate_cluster	0.771	0.418-1.421	0.386	220	22
GEE Poisson, exchangeable	gee_poisson_exchangeable	0.876	0.519-1.476	0.602	220	22
PPML plus ALN-system intensity break	ppml_cluster	1.051	0.839-1.316	0.652	220	22

ADDM PCV7

Specification	Estimator	RR	95% CI	p	N	Clusters
Site and cohort fixed effects	log-prevalence OLS with site-cluster inference	1.072	0.953-1.206	0.190	48	6
Intensity-specific linear trend	log-prevalence OLS with site-cluster inference	0.947	0.822-1.091	0.368	48	6
Intensity-specific flexible spline	log-prevalence OLS with site-cluster inference	1.053	0.971-1.143	0.163	48	6

Notes: Confidence intervals use cluster-robust inference with cluster-minus-one t critical values where applicable. Effect units remain design-specific and are not pooled.

S4. Placebo timing, future exposure, and randomization-style inference

The analysis contains nineteen placebo diagnostics. Randomization tests use 1,999 deterministic permutations. Permuting state PCV7 intensity yields p=0.333; permuting SchoolVaxView MMR coverage across states within cohort yields p=0.251. The PCV7 joint lead test gives p=0.636, SchoolVaxView future coverage gives p=0.287, and the six-site ADDM intensity permutation gives p=0.409. Four English pre-program cutoffs are null (minimum p=0.104). Welsh authority-uptake permutations, the Northern Ireland cutoff rank, and the Yokohama false-date exercises are also non-rejecting.

Two of these nineteen placebo diagnostics fail: Spain's regional varicella event study rejects the joint pretrend null (p=0.020), and Wales rejects its dynamic pretrend null (p=0.010). These failures are not averaged away. They are recorded as design-level identification failures and prevent causal interpretation even though other placebo exercises in those panels are non-rejecting. The separate outcome-specific Scottish tests in S57 also reject for hearing support and, in the fully reported sample, speech/language support.

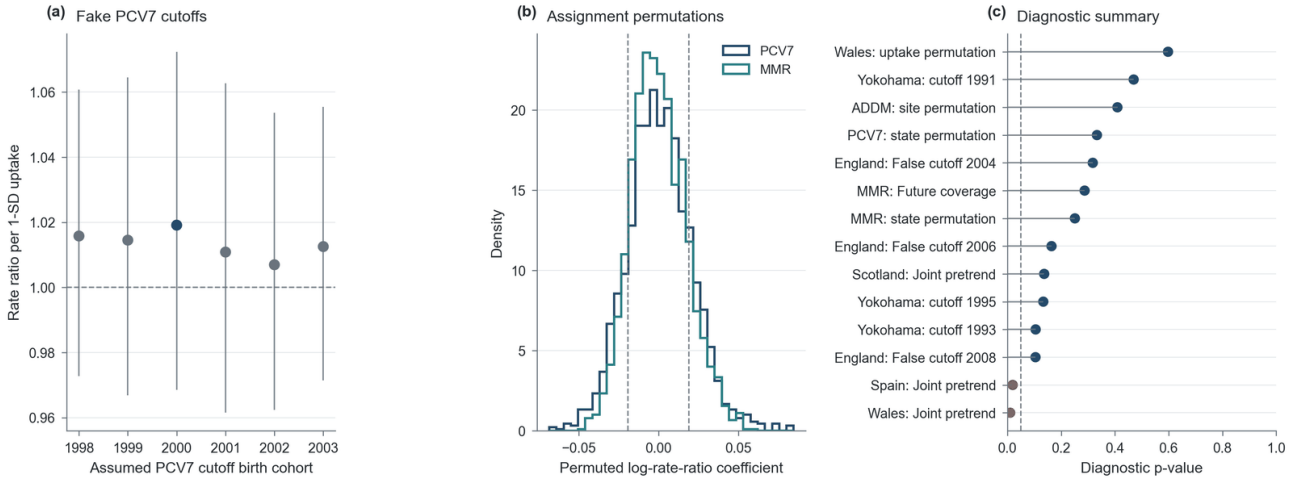


Figure S3. Fake-event, assignment-permutation, and pretrend diagnostics

Table S3. Placebo and pretrend diagnostics

Design	Diagnostic	Family	p	Result	Details
U.S. IDEA PCV7	State-intensity permutation	assignment placebo	0.333	Pass	1999 permutations of state uptake intensity; state and cohort FE retained.
U.S. IDEA PCV7	Joint event-study leads	pretrend placebo	0.636	Pass	All non-reference pre-PCV7 cohort interactions.
SchoolVaxView MMR	Within-cohort state permutation	assignment placebo	0.251	Pass	1999 permutations of MMR coverage across states within birth cohort.
SchoolVaxView MMR	Future coverage	future-exposure placebo	0.287	Pass	Next cohort's kindergarten coverage substituted for contemporaneous coverage.
ADDM PCV7	Site-intensity permutation	assignment placebo	0.409	Pass	Exact/exhaustive reassignment across the six recurring sites.
England rotavirus	False cutoff 2004	timing placebo	0.318	Pass	Cutoff restricted to pre-program birth cohorts.
England rotavirus	False cutoff 2006	timing placebo	0.164	Pass	Cutoff restricted to pre-program birth cohorts.
England rotavirus	False cutoff 2008	timing placebo	0.104	Pass	Cutoff restricted to pre-program birth cohorts.
England rotavirus	False cutoff 2010	timing placebo	0.922	Pass	Cutoff restricted to pre-program birth cohorts.
Wales rotavirus	Authority-uptake permutation	assignment placebo	0.598	Pass	399 local-authority uptake-profile permutations.
Wales rotavirus	Interaction permutation	assignment placebo	0.925	Pass	399 local-authority uptake-profile permutations.
Northern Ireland rotavirus	Placebo-cutoff rank	timing placebo	0.875	Pass	True cutoff ranked against all feasible hypothetical cohort cutoffs.
Yokohama MMR withdrawal	Withdrawal cutoff 1990	timing placebo	0.739	Pass	Actual 1993 and false withdrawal dates reported together.
Yokohama MMR withdrawal	Withdrawal cutoff 1991	timing placebo	0.470	Pass	Actual 1993 and false withdrawal dates reported together.
Yokohama MMR withdrawal	Withdrawal cutoff 1993	timing placebo	0.105	Pass	Actual 1993 and false withdrawal dates reported together.
Yokohama MMR withdrawal	Withdrawal cutoff 1995	timing placebo	0.133	Pass	Actual 1993 and false withdrawal dates reported together.
Scotland rotavirus	Joint pretrend	pretrend placebo	0.137	Pass	Dynamic pre-exposure coefficients.
Wales rotavirus	Joint pretrend	pretrend placebo	0.010	Fail	Fails before the ALN outcome-system transition.
Spain varicella	Joint pretrend	pretrend placebo	0.020	Fail	Regional convergence design fails the joint lead test.

Notes: A pass means the diagnostic does not reject its placebo null at five percent. It is not evidence that all identifying assumptions hold.

S5. Claim-relevant comparator conditions

The cross-condition audit estimates the preferred specification for eight U.S. IDEA categories, six Scottish additional-support outcomes, six Welsh ALN outcomes, and four Swedish specialist-care diagnosis groups. Autism is displayed as the target outcome; all other rows are claim-relevant comparator outcomes that can also reveal shared reporting, service, cohort, or area trends. None is assumed unaffected by vaccination.

No cross-condition estimate rejects after Benjamini-Hochberg correction within design. In the U.S. IDEA panel, autism is RR 1.013 (95% CI 0.981 to 1.045); speech/language is RR 0.995 (0.957 to 1.034); intellectual disability is RR 1.028 (0.968 to 1.090); and the other categories are likewise imprecise or close to one after flexible intensity trends. Scotland and Wales have much wider intervals, so their comparator nulls are weak tests rather than strong evidence of specificity. In Sweden, the preferred F90 estimate is nominally inverse (RR 0.742, p=0.019) but does not survive the four-diagnosis correction (q=0.074); F84, F80, and F82 do not reject.

These comparisons test whether the exposure measure is proxying a broad change in administrative identification. They do not identify whether vaccination biologically causes or prevents each comparator condition. Scottish hearing support is RR 0.566 (0.233-1.373; p=0.200) under the preferred propensity specification, with its own pretrend rejection. S57 gives source-token checks, separate outcome-specific event studies and count-model sensitivities for all six Scottish support outcomes.

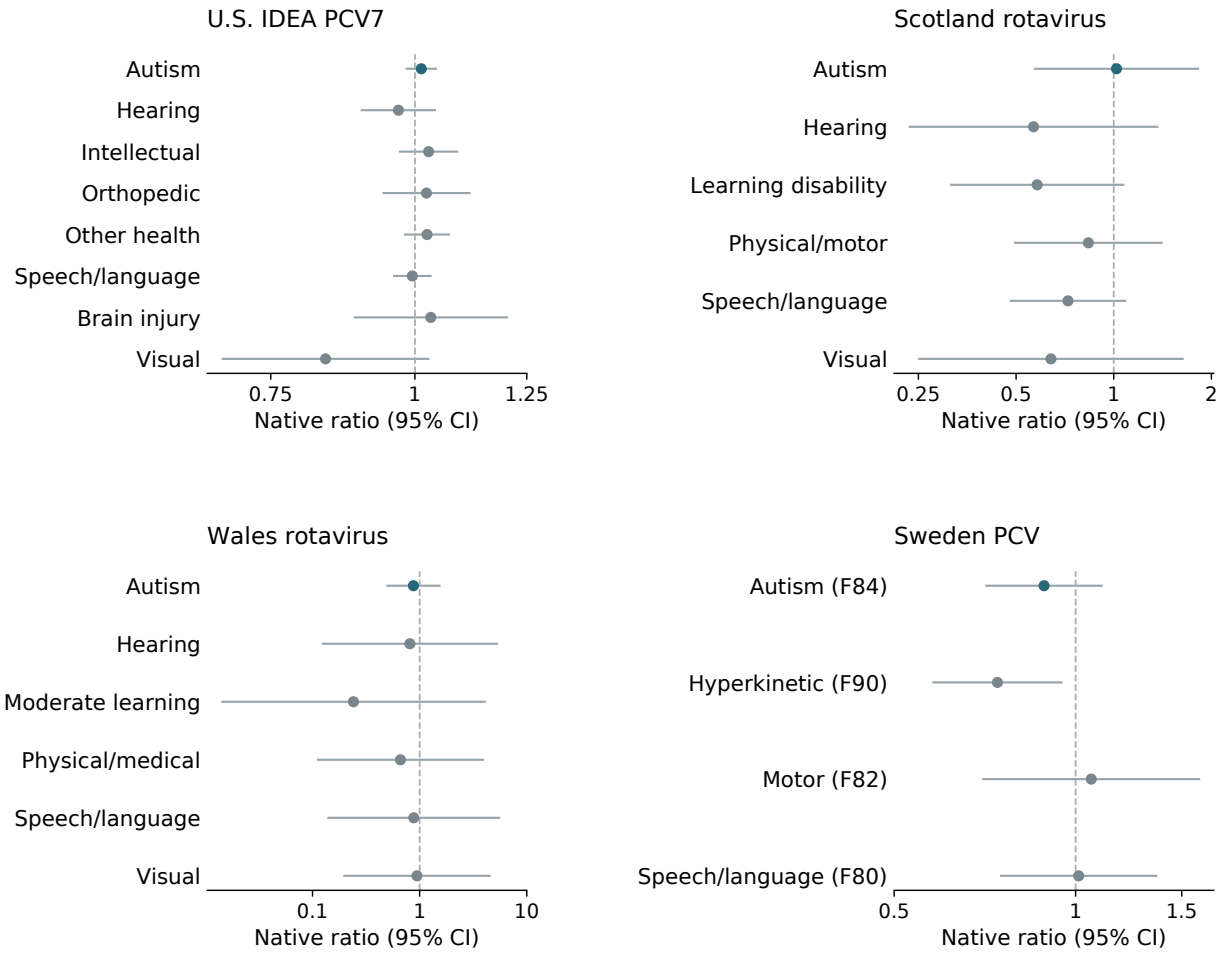


Figure S4. Autism and comparator-condition estimates under preferred specifications

Table S4. Claim-relevant cross-condition outcomes

Design	Condition	RR	95% CI	p	BH q	Target?
U.S. IDEA PCV7	Autism	1.013	0.981-1.045	0.419	0.671	Autism
U.S. IDEA PCV7	Hearing impairments	0.968	0.898-1.043	0.381	0.671	Comparator
U.S. IDEA PCV7	Visual impairments	0.837	0.680-1.029	0.090	0.671	Comparator
U.S. IDEA PCV7	Orthopedic impairments	1.023	0.937-1.117	0.602	0.777	Comparator
U.S. IDEA PCV7	Traumatic brain injury	1.032	0.885-1.204	0.680	0.777	Comparator
U.S. IDEA PCV7	Speech or language impairments	0.995	0.957-1.034	0.785	0.785	Comparator
U.S. IDEA PCV7	Intellectual disability	1.028	0.968-1.090	0.360	0.671	Comparator
U.S. IDEA PCV7	Other health impairments	1.025	0.979-1.073	0.294	0.671	Comparator
Scotland rotavirus	Autism	1.020	0.567-1.833	0.946	0.946	Autism
Scotland rotavirus	Hearing	0.566	0.233-1.373	0.200	0.399	Comparator
Scotland rotavirus	Visual	0.641	0.250-1.643	0.342	0.514	Comparator
Scotland rotavirus	Physical Motor	0.835	0.493-1.416	0.492	0.591	Comparator
Scotland rotavirus	Speech Language	0.723	0.478-1.093	0.119	0.358	Comparator
Scotland rotavirus	Learning Disability	0.581	0.313-1.078	0.083	0.358	Comparator
Wales rotavirus	Autism	0.875	0.489-1.566	0.639	0.940	Autism
Wales rotavirus	Physical Medical	0.662	0.110-3.981	0.638	0.940	Comparator
Wales rotavirus	Hearing	0.811	0.122-5.396	0.820	0.940	Comparator
Wales rotavirus	Visual	0.944	0.194-4.601	0.940	0.940	Comparator
Wales rotavirus	Speech Language	0.881	0.137-5.648	0.888	0.940	Comparator
Wales rotavirus	Moderate Learning	0.241	0.014-4.141	0.310	0.940	Comparator
Sweden PCV	Autism category (F84)	0.886	0.708-1.109	0.292	0.584	Autism
Sweden PCV	Speech/language (F80)	1.012	0.749-1.366	0.939	0.939	Comparator
Sweden PCV	Motor development (F82)	1.062	0.700-1.610	0.778	0.939	Comparator
Sweden PCV	Hyperkinetic disorders (F90)	0.742	0.578-0.951	0.019	0.074	Comparator

Notes: Benjamini-Hochberg q-values are calculated within design across the displayed conditions. Other outcomes are not assumed unaffected by vaccination. Scottish counts use all three school sectors; outcome-specific pretrend and suppression checks are in S57. These are not individually identified biological effects.

S6. Influence, small-cluster, source, and measurement robustness

Influence and inference checks include U.S. state-intensity permutations; the complete 51-state delete-one distribution and cluster-jackknife standard error; alternative NIS intensity-year windows and a rank transformation; ADDM exact wild-cluster and exhaustive site-intensity tests; leave-one-site-out ADDM estimates; English leave-one-cohort ranges and wild-cluster intervals; Welsh suppression bounds, ALN-break adjustments, authority permutations, and leave-one-authority estimates; Northern Ireland leave-one-report and placebo-cutoff analyses; Swedish county bootstrap inference and legacy-mapped-birth sensitivity; and Spain donor/pre-fit diagnostics.

IDEA's preferred RR is 1.013. Omitting one state at a time produces RRs from 1.004 to 1.020; the largest absolute delete-one change is 0.55 reference-model standard errors. The delete-one cluster-jackknife standard error is 0.0150, compared with 0.0156 conventionally. Rebuilding early PCV intensity from individual 2001, 2002, or 2003 NIS waves and leave-one-year-out windows gives RRs from 1.003 to 1.019; a standardized rank exposure gives 1.011. A centered quadratic post-intensity term does not reject linearity (p=0.117). Clustering by nine Census divisions gives SE 0.0159 and p=0.447; deleting one division at a time gives RRs from 0.999 to 1.021, with a maximum absolute change of 0.90 reference-model standard errors. These checks do not reveal concentration in one state, region, proxy year, or simple linear exposure scale. Early-intensity states nevertheless have higher preperiod outcome levels (RR 1.242 per intensity SD, HC3 p<0.001). State fixed effects absorb permanent level differences, but this imbalance reinforces the need to rely on within-state cohort changes and pretrend diagnostics rather than cross-sectional levels.

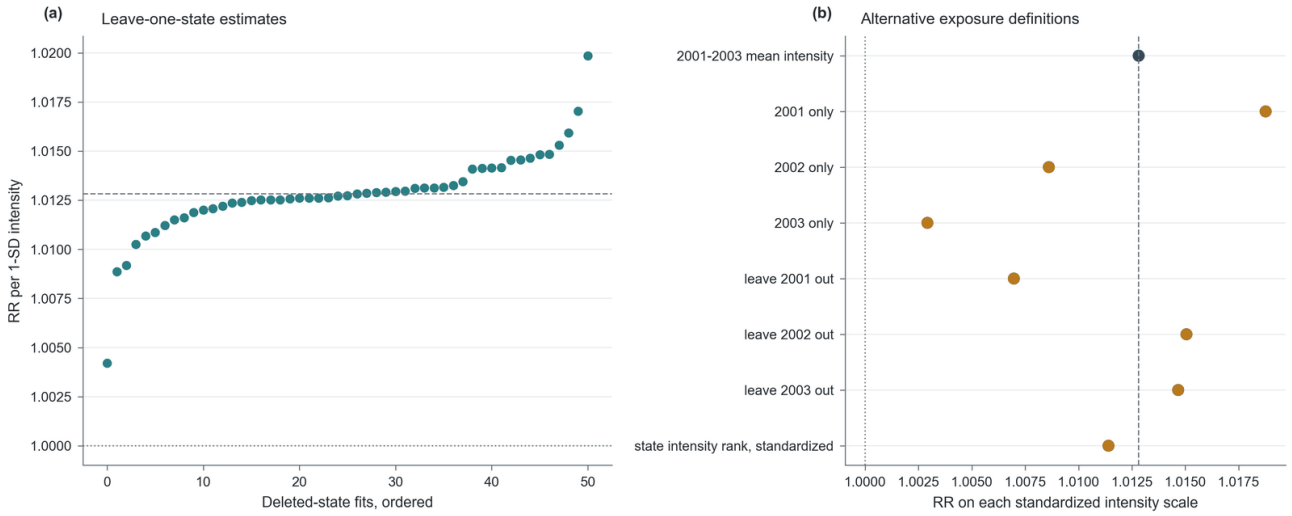


Figure S9. IDEA PCV7 state-deletion and exposure-definition sensitivity

Table S9. IDEA PCV7 influence and exposure-definition audit

Diagnostic	Result	Inference / range	Interpretation
Preferred model	RR 1.013	SE 0.0156; p=0.419	2001-2003 mean intensity
Delete-one state	RR 1.004-1.020	max |DFBETA| 0.55	No single-state concentration
Cluster jackknife	RR 1.013	SE 0.0150; p=0.400	Nearly identical to conventional clustered SE
Alternative NIS years	RR 1.003-1.019	six constructions	Proxy-year sensitivity
Rank intensity	RR 1.011	SE 0.0150	Monotone functional-form sensitivity
Quadratic intensity	exp(beta2) 1.018	p=0.117	Centered quadratic term; not the main RR
Spatial dependence	division-cluster SE 0.0159	leave-division RR 0.999-1.021	Nine-division low-df sensitivity
Preperiod level balance	RR 1.242	HC3 p=3.33e-07	Permanent level imbalance; controlled by state FE

Notes: Delete-one fits are influence diagnostics, not separate hypothesis tests. Alternative standardized exposure scales are not numerically identical causal units.

These tests distinguish two kinds of instability. Sampling instability is visible in ADDM, Scotland, and Wales through broad intervals and leave-one-cluster movement. Structural instability is visible in England, Wales, Northern Ireland, Sweden, and Spain when cohort trends, APC normalization, outcome-system changes, or pretrends alter the estimate. Structural failures receive greater weight than an isolated favorable coefficient.

S7. Complete design-by-test atlas

Filled circles in the atlas indicate an implemented, adequate check; open circles indicate a limited check; crosses indicate a failed placebo or pretrend. These symbols record whether a test was implemented and what it found. Separate columns report counterfactual credibility and overfit risk, preventing a row of implemented checks from being mistaken for a credible design. The matrix is not a score to be summed. The U.S. IDEA PCV7 family combines the best-measured ecological rollout intensity, two estimated exact-age leads plus an omitted reference, claim-relevant cross-outcome comparisons, estimator checks, exchangeability-dependent permutation diagnostics, and adequate moderate-effect power. Its longer historical diagnostic crosses an outcome-source transition and is not counted as five clean leads. Other panels contribute distinct outcome systems or useful first stages but remain limited by short preperiods, outcome mismatch, national rollouts, small clusters, or system breaks.

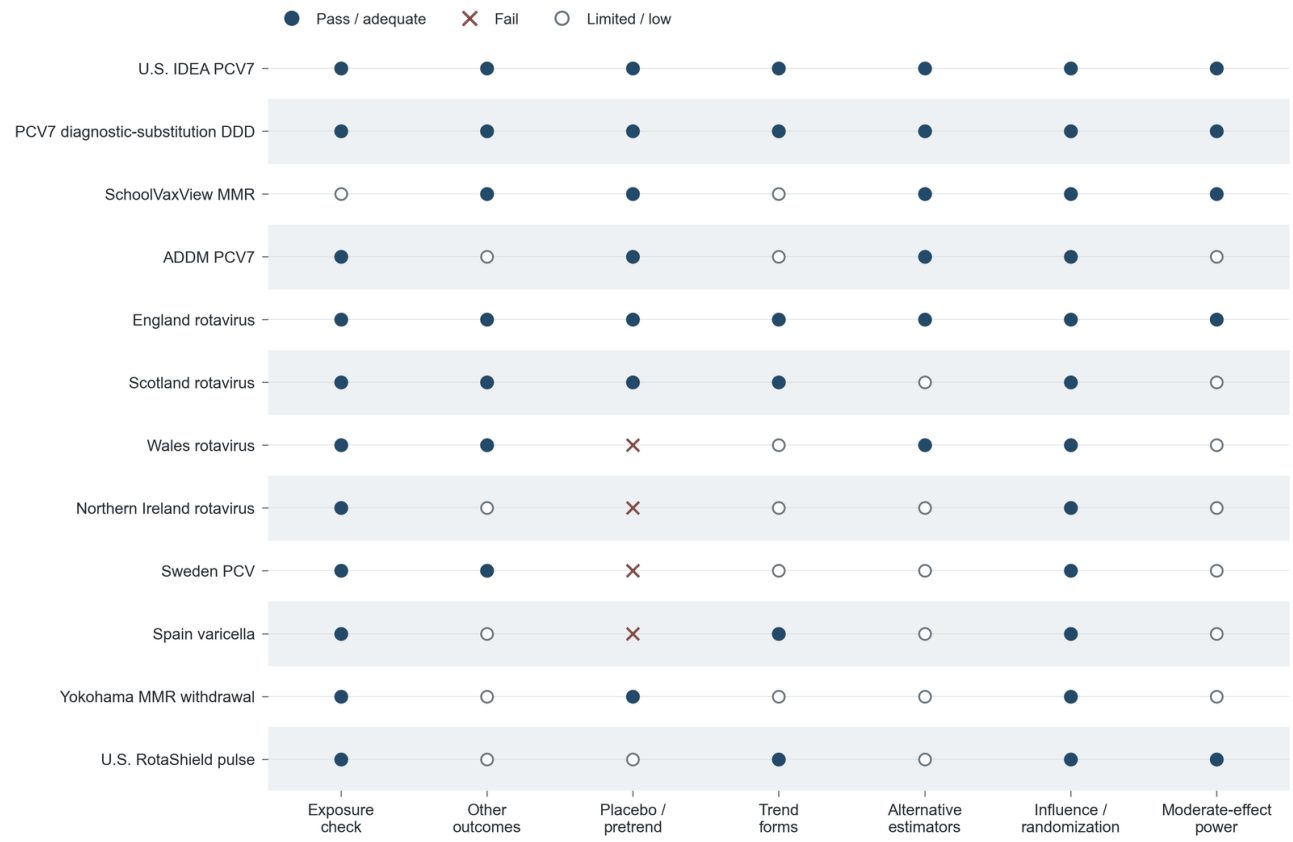


Figure S5. Complete design-by-test robustness atlas

Table S5. Design-by-test robustness atlas

Design	Exp.	Outc.	Plac.	Trend	Est.	Infl.	Power	Counterfactual	Overfit	Interpretation
U.S. IDEA PCV7	pass	pass	pass	pass	pass	pass	80%+	Credible (ecol.)	low–mod.	Most complete ecological design; only three preperiods
PCV7 diagnostic-substitution DDD	pass	pass	pass	pass	pass	pass	80%+	Credible falsif.	low	Strong diagnostic-substitution falsification; not a person-level effect
SchoolVaxView MMR	limited	pass	pass	limited	pass	pass	80%+	limited	high	Only three outcome cohorts; state trends are near-saturated
ADDM PCV7	pass	limited	pass	limited	pass	pass	<80%	fragile	high	Six-site small-cluster sensitivity panel; trend estimates move materially
England rotavirus	pass	pass	pass	pass	pass	pass	80%+	Trend-sensitive	mod./high	Estimator-stable but APC/cohort-sensitive national step
Scotland rotavirus	pass	pass	pass	pass	limited	pass	Low	Weak	moderate	Eight preperiods; proxy uptake and wide intervals
Wales rotavirus	pass	pass	fail	limited	pass	pass	Low	failed	high	Pretrend and ALN break fail; two preperiods
Northern Ireland rotavirus	pass	limited	fail	limited	limited	pass	Low	failed	high	APC/placebo instability
Sweden PCV	pass	pass	fail	limited	limited	pass	Low	failed	mod.–high	Outcome and pretrend mismatch
Spain varicella	pass	limited	fail	pass	limited	pass	Low	failed	low/high bias	Joint pretrend rejects; curvature selected in preperiod
Yokohama MMR withdrawal	pass	limited	pass	limited	limited	pass	Low	limited	high	Nine aggregate cohorts
U.S. RotaShield pulse	pass	limited	limited	pass	limited	pass	80%+	limited	moderate	Only one clean lead

Notes: Pass, limited, and fail in the test columns describe implementation or the diagnostic result; 80%+ marks conditional power of at least 80 percent for the selected moderate effect. Counterfactual credibility and overfit risk are separate judgments. A failed pretrend is not offset by other checks.

S8. Interpretation and safeguards against specification search

The multiverse should be read in three steps. First, ask whether the event has a verified exposure first stage. Second, ask whether pretrends, placebos, measurement continuity, and comparator outcomes support the identifying assumptions. Only then ask whether the effect estimate is stable across estimators and trend forms. Statistical power is reported last because high power cannot rescue a biased counterfactual.

The complete catalog records every test's identifier, design, family, statistic, p-value where applicable, status, and interpretation. IDEA PCV7 and its diagnostic-substitution DDD retain their model-native estimates, but the exposure audit prevents treating their original individual-risk mapping as confirmed. SchoolVaxView, ADDM, other countries, schedule interactions, and cross-condition outcomes supply distinct or exploratory contrasts. The U.S. PCV7 exercises share exposure timing and are not independent replications. Benjamini-Hochberg corrections are applied within coherent outcome or interaction families; no global q-value or omnibus count combines heterogeneous estimands. Passing a diagnostic is not proof of identification.

Modern staggered-adoption estimators are not added ritualistically. The headline designs use common-timing continuous intensities or national programme steps rather than staggered binary adoption with a clean not-yet-treated group. Synthetic control is likewise not used for the continuous-intensity IDEA panel. Additional APC flexibility, machine-learning nuisance fits, or pooled meta-estimates would add modeling choices without new identifying variation.

S9. Individual-equivalent RR and OR calibration

Individual-scale effects can be estimated from aggregate contrasts under explicit identification and aggregation assumptions. The crucial requirement is that the fitted contrast match the stipulated exposure mixture. A raw pair of coverage shares cannot automatically convert a trend-adjusted rollout coefficient.

For an exact comparison of exposure shares p0 and p1 and a stable affected fraction s of baseline outcome risk, the homogeneous-mixture ratio is [1+p1*s*(R-1)]/[1+p0*s*(R-1)]. The fraction s is not generally the population fraction eligible for vaccination. The formula supports introductions and withdrawals, with reversed ordering when p1<p0. Confidence-set inversion retains empty, unbounded and zero-contrast cases rather than clipping them to plausible-looking estimates.

The archived five-design table uses stipulated coverage pairs for IDEA PCV7, IDEA RotaShield, SchoolVaxView MMR, ADDM PCV7 and ADDM RotaShield. Those pairs are not certified as estimator-matched first stages. PCV7's early-uptake ranking is specifically inconsistent with the positive two-share interpretation (S52); ADDM shares that exposure problem despite a different outcome system. The other nominal gradients require their own design-weighted forward maps. They are not converted into empirical individual-risk bounds by assertion.

Current direct mixture estimates and signed-score inversions appear in S53, including Japan. Native estimates remain available for every other design. RR-to-RD or RR-to-OR conversion additionally requires the unexposed risk for the same population and time horizon: RD=p0*(R-1) and OR=R*(1-p0)/(1-R*p0), where p0 here denotes baseline outcome risk, not vaccination share. Values implying exposed risk above one are infeasible, not infinite odds estimates. A hazard ratio requires a survival model before conversion to a cumulative risk ratio.

S10. Pretrend adjustment and one-step-up trend sensitivity

The pretrend audit complements joint lead tests with coefficient stability across baseline fixed effects, intensity-specific linear trends, flexible splines, outcome-system breaks, and a two-stage preperiod-only detrending estimator. The latter estimates a linear differential trend using only unexposed cohorts, extrapolates that trend into the postperiod, and requests 1,999 cluster-pairs bootstrap draws to propagate first-stage uncertainty. In ADDM, 43 draws contain fewer than three distinct resampled sites and are discarded by rule, leaving 1,956 valid draws; the other estimated panels retain all 1,999. It is estimated only when at least three preperiods are available. Wales has two and is reported as not estimable.

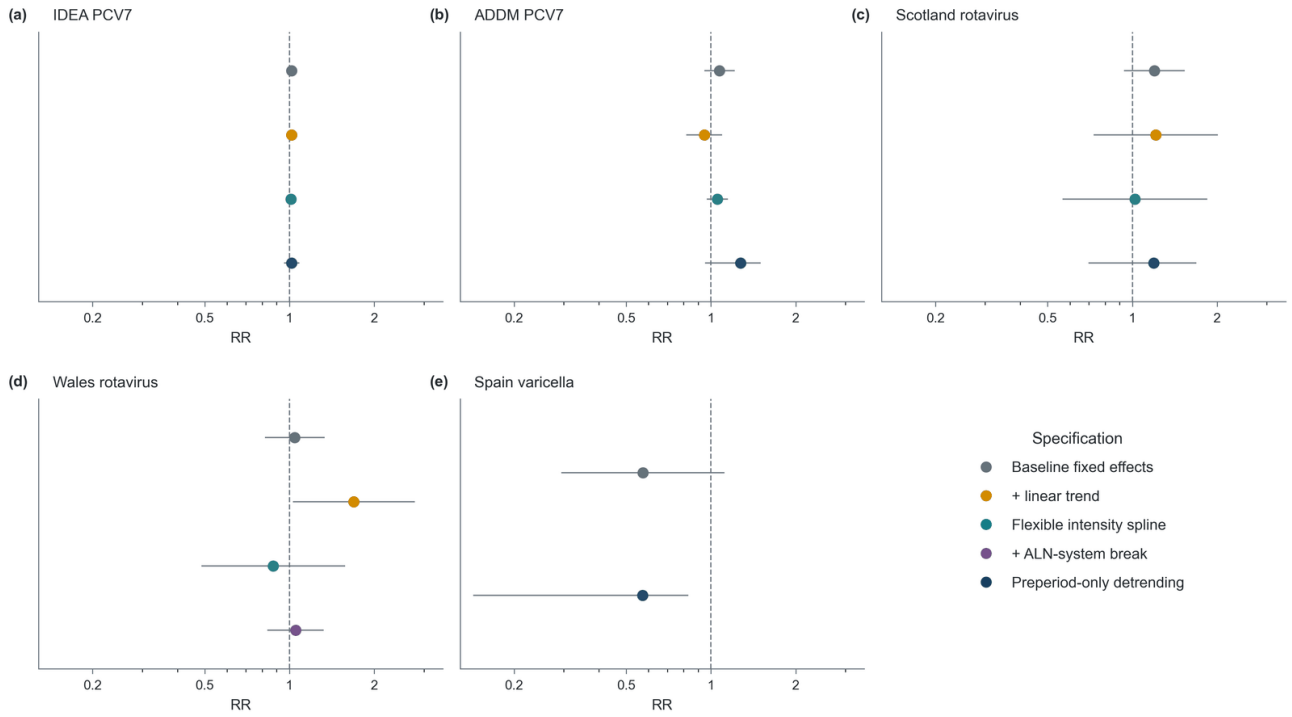


Figure S8. Baseline, one-step-up, flexible-trend, and preperiod-only detrended estimates

Table S8. Trend-adjusted and preperiod-only detrended estimates

Design	Specification	RR	95% interval	Preperiods	Status
IDEA PCV7	two-way fixed effects	1.017	0.979-1.057	-	existing fitted specification
IDEA PCV7	one-step-up linear trend	1.016	0.987-1.047	-	existing fitted specification
IDEA PCV7	flexible intensity spline	1.013	0.981-1.045	-	existing fitted specification
ADDM PCV7	two-way fixed effects	1.072	0.953-1.206	-	existing fitted specification
ADDM PCV7	one-step-up linear trend	0.947	0.822-1.091	-	existing fitted specification
ADDM PCV7	flexible intensity spline	1.053	0.971-1.143	-	existing fitted specification
Scotland rotavirus	two-way fixed effects	1.195	0.938-1.524	-	existing fitted specification
Scotland rotavirus	one-step-up linear trend	1.209	0.731-1.999	-	existing fitted specification
Scotland rotavirus	flexible intensity spline	1.020	0.567-1.833	-	existing fitted specification
Wales rotavirus	two-way fixed effects	1.044	0.822-1.325	-	existing fitted specification
Wales rotavirus	one-step-up linear trend	1.692	1.033-2.771	-	existing fitted specification
Wales rotavirus	flexible intensity spline	0.875	0.489-1.566	-	existing fitted specification
Wales rotavirus	spline plus ALN-system break	1.051	0.839-1.316	-	existing fitted specification
IDEA PCV7	preperiod-only linear detrending	1.016	0.962-1.077	3	estimated; pairs cluster bootstrap propagates first-stage trend uncertainty
ADDM PCV7	preperiod-only linear detrending	1.275	0.958-1.494	3	estimated; pairs cluster bootstrap propagates first-stage trend uncertainty
Scotland rotavirus	preperiod-only linear detrending	1.188	0.702-1.676	8	estimated; pairs cluster bootstrap propagates first-stage trend uncertainty
Wales rotavirus	preperiod-only linear detrending	not estimable	not estimable	2	not estimable: fewer than three preperiods
Spain varicella	two-way fixed effects	0.573	0.296-1.112	-	existing fitted specification
Spain varicella	preperiod-only linear detrending	0.573	0.144-0.826	7	estimated; pairs cluster bootstrap propagates first-stage trend uncertainty

Notes: Preperiod-only detrending requests 1,999 cluster-pairs bootstrap draws and is not estimated with fewer than three preperiods. ADDM retains 1,956 valid draws after excluding resamples with fewer than three distinct sites. Model families remain design-specific.

IDEA PCV is stable across all variants. ADDM is substantially more sensitive and has only six clusters. Scotland remains too imprecise to distinguish moderate effects. Wales changes sign and magnitude across trend and ALN-system

specifications. Spain's inverse point estimate persists after linear preperiod extrapolation, but its rejected event-study leads continue to disqualify a causal interpretation. A flexible trend or detrending exercise cannot repair a visibly invalid counterfactual.

S11. Model complexity, leverage, and preperiod-only selection

The model-geometry audit reports observations, independent clusters, periods, preperiods, design-matrix rank, residual degrees of freedom, leverage, and outcome-weight concentration for the principal flexible specifications. The preferred IDEA model uses 353 observations, 51 state clusters, matrix rank 60, and 293 residual degrees of freedom. Its largest state has 12.4 percent of outcome weight, and the inverse-Herfindahl effective cluster count by weight is 22.0. This is not evidence that all identifying assumptions hold, but it is inconsistent with a model mechanically saturated by its nuisance terms.

The short panels are different. SchoolVaxView's state/cohort model has rank 53 in 142 observations, while its state-specific trend sensitivity has rank 101 and only 41 residual degrees of freedom. ADDM's flexible model has rank 16 in 48 observations but only six independent sites and high cluster leverage. Wales has 34 fitted dimensions and only two preperiods. England has ample residual degrees of freedom, yet remains counterfactual-fragile because a national programme step is inseparable from cohort/APC restrictions. Spain has relatively light parameter burden; its failure is bias revealed by seven preperiods, not computational overfit.

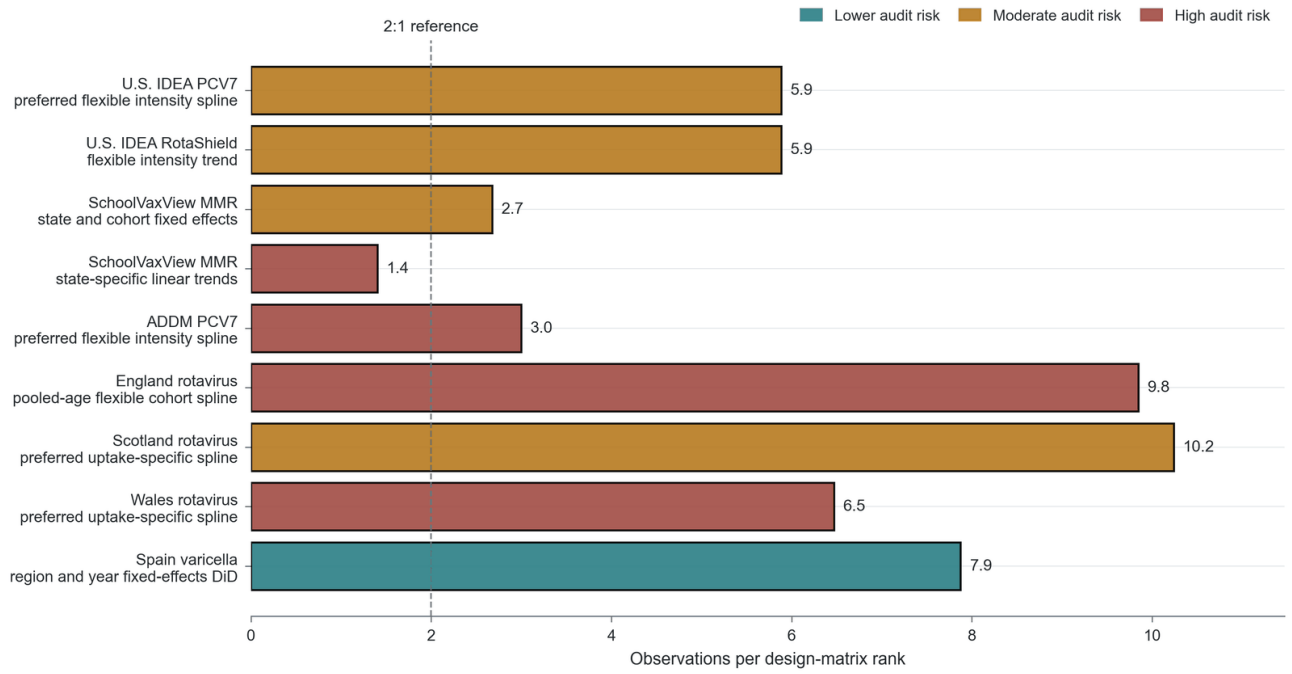


Figure S10. Model flexibility relative to observed support

Table S10. Model rank, cluster support, and leverage

Design	Specification	N / G / pre	Rank	Residual df	N/rank	Max weight	Overfit / identification
U.S. IDEA PCV7	preferred flexible intensity spline	353 / 51 / 3	60	293	5.88	12.4%	low to moderate; best-supported ecological design
U.S. IDEA RotaShield	flexible intensity trend	353 / 51 / 1	60	293	5.88	12.4%	moderate; thin timing support
SchoolVaxView MMR	state and cohort fixed effects	142 / 50 / 0	53	89	2.68	12.0%	moderate; three-cohort dose-response; no event-study leads
SchoolVaxView MMR	state-specific linear trends	142 / 50 / 0	101	41	1.41	12.0%	high / near-saturated; sensitivity only
ADDM PCV7	preferred flexible intensity spline	48 / 6 / 3	16	32	3.00	16.7%	high; small-cluster sensitivity panel
England rotavirus	pooled-age flexible cohort spline	128 / 23 / 15	13	115	9.85	6.8%	moderate parameter burden / high counterfactual fragility; estimator-stable, trend-form fragile
Scotland rotavirus	preferred uptake-specific spline	512 / 32 / 8	50	462	10.24	9.9%	moderate; long preperiod, low precision
Wales rotavirus	preferred uptake-specific spline	220 / 22 / 2	34	186	6.47	12.0%	high for counterfactual selection; failed pretrend and ALN system break
Spain varicella	region and year fixed-effects DiD	252 / 18 / 7	32	220	7.88	27.6%	low parameter burden; failed differential pretrend

Notes: G is the independent clustering dimension and pre is the number of preperiods. N/rank is descriptive, not a universal overfit cutoff; identification can fail even with ample residual degrees of freedom.

Table S11. Preperiod-only differential-path selection

Design	Preperiods	Selected path	LOOCV RMSE	Status
U.S. IDEA PCV7	3	constant differential gap	0.0019	selected
ADDM PCV7	3	linear differential trend	0.0069	selected
Scotland rotavirus	8	constant differential gap	0.0140	selected
Wales rotavirus	2	not estimable	-	fewer than three preperiods
Spain varicella	7	quadratic differential trend	0.1760	selected

Notes: Leave-one-preperiod-out prediction is performed on period-specific exposure-gradient estimates. With three preperiods only constant and linear paths are compared; Wales is not estimable.

A preperiod-only trajectory cross-validation compares constant, linear, and, with at least five preperiods, quadratic differential paths. It selects a constant gap for IDEA and Scotland, a linear path for ADDM, and a quadratic path for Spain. Wales cannot be selected with two preperiods. These results explain rather than replace the full event-study evidence: IDEA does not require added preperiod curvature, ADDM's preferred spline is not supported as a clean confirmatory model, and Spain's counterfactual visibly drifts before policy convergence.

S12. Simultaneous event-study bands and formal sensitivity scope

Pointwise event-study intervals are supplemented with conservative Bonferroni t bands that provide 95 percent familywise coverage across each path's non-reference coefficients. The simultaneous bands remain centered near zero for IDEA PCV7 and the diagnostic-substitution DDD. They widen sharply for ADDM, Scotland, Wales, and RotaShield, making the limits of short or small-cluster panels visually explicit. Familywise bands do not repair the Welsh failed lead or establish a RotaShield multi-period pretrend.

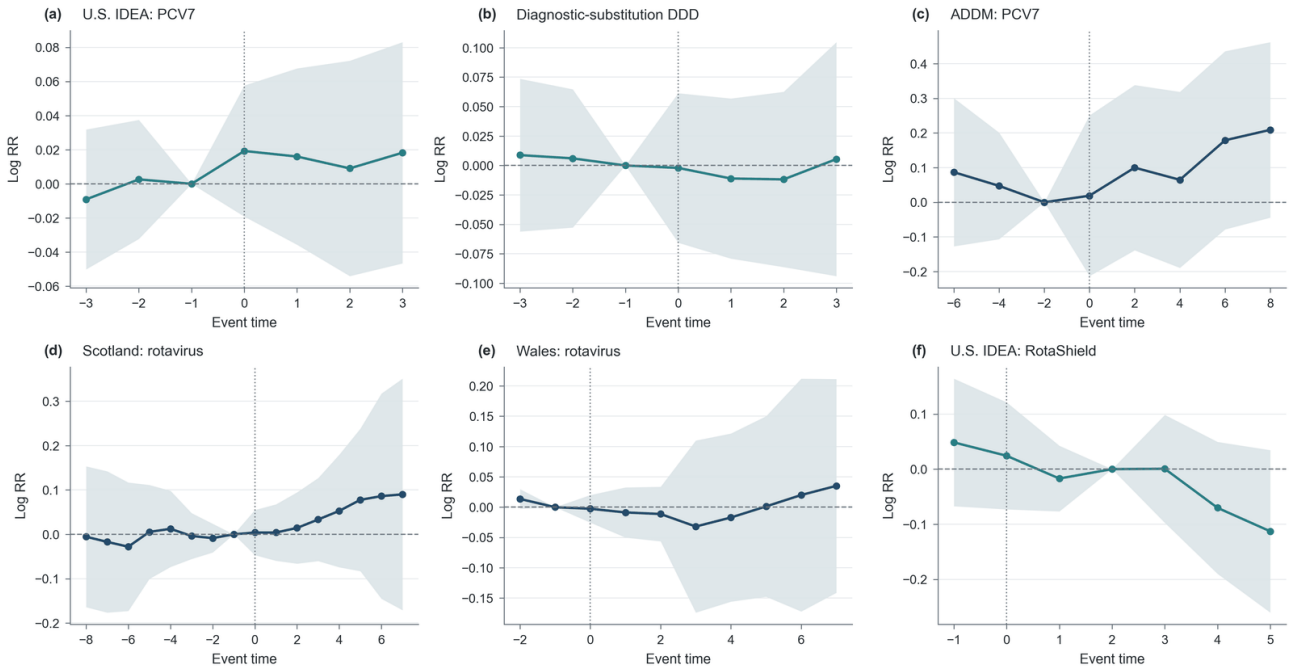


Figure S11. Event-study paths with familywise 95 percent confidence bands

Rambachan-Roth sensitivity bounds are not relabeled or approximated. The official HonestDiD implementation (version 0.2.6) produced the archived flagship U.S. IDEA PCV7 and long-lead sensitivity outputs in files 171 and 177; it is an optional R sensitivity rather than a step in the default offline public runner. Section 5.1 of the main paper reports the key bounds. The relative-magnitudes bound on the average post-period effect excludes the specified moderate effect RR 1.10 through M-bar 0.9 and fails from M-bar 1.0; the first post-period cohort taken alone excludes it across the whole grid to M-bar 2. Two limitations are stated rather than smoothed: several headline contrasts are common-timing continuous-intensity designs with only two non-reference leads, so the benchmark pre-period violation may be small by chance - which makes the exercise uninformative rather than invalid, since the conditional-least-favorable hybrid remains valid - and the long-lead panel's own bound is honestly uninformative, reaching RR 1.10 already at M-bar 0.1 because a noisy first recovered lead sets the benchmark and fourteen post periods accumulate permitted violations. Estimator-matched noncentral-Wald injection confirms that the registered two-lead diagnostic detects only 20.5 percent of a one-percent annual differential cohort trend and 65.9 percent of a two-percent trend (files 2884-2885). Alongside the formal bounds the paper continues to report event coefficient paths, joint lead tests, one-step-up estimates, preperiod-only model selection, bootstrap propagation, and simultaneous bands.

S13. Power, equivalence, and Bayesian calibration

Power, equivalence, and Bayes factors are reported only after identification diagnostics. File 103 gives the numeric 80 percent MDE and power at RR 1.10 for every integrated estimand; file 107 reports whether each confidence interval lies within plus or minus 10 or 20 percent; and file 125 reports approximate normal-likelihood Bayes factors for selected primary designs under positive half-normal alternatives scaled to RR 1.10 and 1.25 on the design-specific reduced-form unit. An interval that includes one is not, by itself, evidence of equivalence.

The best-supported IDEA PCV7 estimate has more than 99.9 percent conditional power for RR 1.10 at its one-SD intensity unit, an 80 percent MDE of 1.046, and a 95 percent interval contained within plus or minus 10 percent. Its BF01 ranges from 2.83 to 6.49 across the two stated positive alternatives. The diagnostic-substitution DDD has 99.0 percent power and BF01 5.24 to 12.06 but remains a differential-outcome falsification estimand. SchoolVaxView's short state/cohort model has 86.3 percent power at RR 1.10, while its state-trend sensitivity has only 27.9 percent. ADDM power against RR 1.10 is 67.3 percent per six-site SD (2.35 points), 10.0 percent per IDEA state SD (8.91 points), and 8.9 percent per ten points, using a two-sided noncentral-t calculation with five degrees of freedom. The corresponding 80 percent MDEs are 1.118, 1.527 and 1.608. The archived normal approximation gives 66.8 percent on the native six-site scale; it is not power against an individual effect. Scotland, Wales, Spain, Sweden, Yokohama, and the ADDM RotaShield pulse are not small-effect designs.

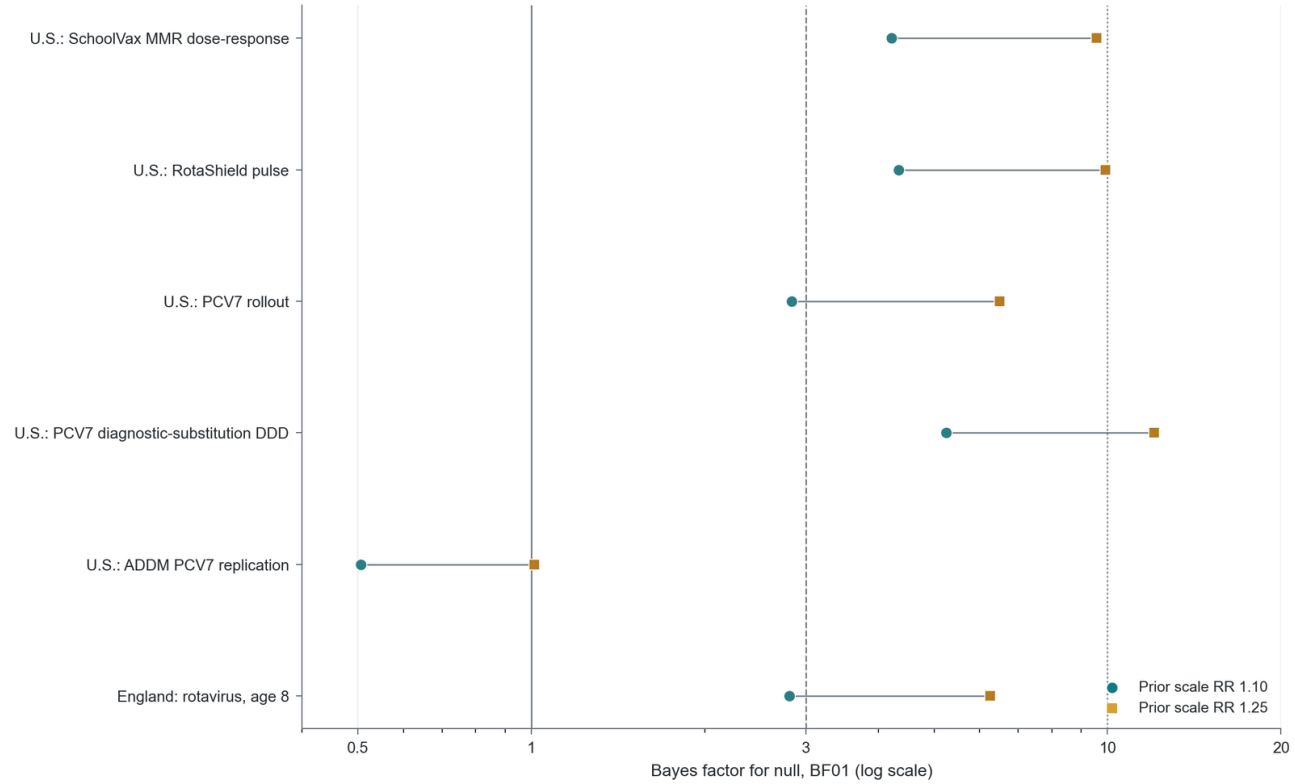


Figure S12. Bayes-factor sensitivity to the positive-alternative scale

Table S12. Design-conditional MDE, power, equivalence, and Bayes calibration

Design	MDE80	Power at RR 1.10	CI within +/-10%	CI within +/-20%	BF01 range	Identification
Japan: MMR coverage	1.619	3.2%	no	no	-	limited: nine aggregate cohorts
Spain: varicella convergence	2.542	5.9%	no	no	-	failed: differential pretrend
U.S.: RotaShield pulse	1.091	87.8%	yes	yes	4.34-9.92	limited: only one clean lead
U.S.: PCV7 rollout	1.046	100.0%	yes	yes	2.83-6.49	best-supported ecological design
U.S.: PCV7 diagnostic-substitution DDD	1.065	99.0%	yes	yes	5.24-12.06	supportive falsification design
U.S.: SchoolVax MMR dose-response	1.091	86.3%	yes	yes	4.21-9.57	limited: only three pandemic-adjacent cohorts
U.S.: ADDM PCV7 replication	1.114	66.8%	no	yes	0.51-1.01	independent but small-cluster clinical-outcome replication
U.S.: ADDM RotaShield replication	1.733	2.5%	no	no	-	clinically closer outcome but severely underpowered
England: rotavirus, pooled ages	1.079	94.2%	no	yes	-	fragile: APC/trend sensitivity
England: rotavirus, age 8	1.097	82.2%	yes	yes	2.81-6.25	fixed-age sensitivity
Scotland: rotavirus ramp	2.291	5.3%	no	no	-	limited: proxy intensity
Wales: rotavirus ramp	2.265	4.9%	no	no	-	failed: pretrend/system break
Northern Ireland: rotavirus	1.265	20.6%	no	no	-	failed: APC/pretrend
Sweden: PCV rollout	1.378	13.3%	no	no	-	failed: pretrend/outcome mismatch

Notes: MDE and power use each model's stated reduced-form unit. Equivalence is descriptive CI containment around one. BF01 uses the approximate normal likelihood and stated positive half-normal priors; no Bayes factor repairs a failed design.

S14. Low-tier and comparator-outcome analyses

The total-evidence extension retains three additional U.S. systems precisely because each illustrates a different failure mode. CDC Medicaid state autism claims yield a naive PCV RR of 1.336 (95% CI 1.115-1.602), but an impossible future schedule share yields RR 1.453 (1.119-1.885) and all fourteen naive fake cohort cutoffs reject. This is a failed administrative design, not positive causal evidence. ACS cognitive difficulty is a null-centered claim-relevant outcome: PCV RR 0.999 (0.942-1.060; MDE80 1.086) and 2006-trio RR 1.011 (0.975-1.049; MDE80 1.052). Precision applies only to cognitive difficulty, not autism. CDC WONDER F84 mortality is both overfit and underpowered; trend-dependent PCV RRs range from 0.056 to 220.95, with enormous leave-one-year instability. Files 245-262 retain all three panels.

The SLAITS-era NSCH replication applies the PCV cohort design to parent-reported individual outcomes rather than IDEA education counts. Across 83,540 children and 1,958 cases, its prevalence ratio is 0.926 per ten points of PCV coverage (95% CI 0.836-1.026; p = 0.141), with an ADHD comparator of 0.949 (0.893-1.008). It is less precise than the IDEA estimate on the same scale, 1.014 (0.980-1.050), and the two are still statistically consistent, but only just: the ratio of the two estimates is 0.913 (0.820-1.017), with an equality p-value of 0.098, or 0.093 after allowing for the NIS exposure data both share. This is the weakest agreement among the six individual-versus-aggregate pairs, whose equality p-values run from 0.10 to 0.93. Its upper limit of 1.026 excludes a 5 percent increase per ten coverage points, but it bounds a state-cohort coverage contrast, not the risk of a vaccinated child relative to an unvaccinated one. Dropping the 2003 wave, whose item asked about autism rather than autism or ASD, gives 0.915 (0.831-1.007); cohorts with directly measured uptake give 0.919 (0.738-1.144); and the MMR and varicella secondaries give 0.785 (0.553-1.114) and 1.035 (0.897-1.193).

Within a survey wave, completed age fixes the assigned birth cohort, making the state, cohort, age and wave effects exactly collinear. The model uses a full-rank basis of that design, preserving the fitted means without inverting a singular Hessian. The public files carry no interview date. Birth cohort is therefore assigned as the wave's first calendar year minus completed age: 2003, 2007 or 2011. This annual convention necessarily misassigns some birth years; about a third of children in every wave are assigned a cohort one year from their true birth year. Assigning the 2011-12 children to the earlier or later adjacent year gives 0.965 (0.839-1.111) and 1.005 (0.903-1.118); the adopted convention gives the lowest of the three estimates, and all three intervals include one. Sources: `161_slaits_pcv7_replication.csv`, `160_concordance_tests.csv`, `163_nis_dependence_adjustment.csv`, `430_small_effect_exclusion.csv` and `qa/slaits_full_rank_20260930/cohort_rule_comparison.json`.

S15. Cross-publication independence and first-stage checks

Texas PEIMS primary-disability reports for 2012-13 through 2025-26 were acquired through the state SAS broker and compared with federal IDEA ages-3-to-21 totals. Across thirteen overlapping years, levels correlate 0.9988 and annual log growth correlates 0.8746. Both publications use related Texas special-education records, so this validates processing and publication but is not an independent sample or epidemiologic replication (files 274-276).

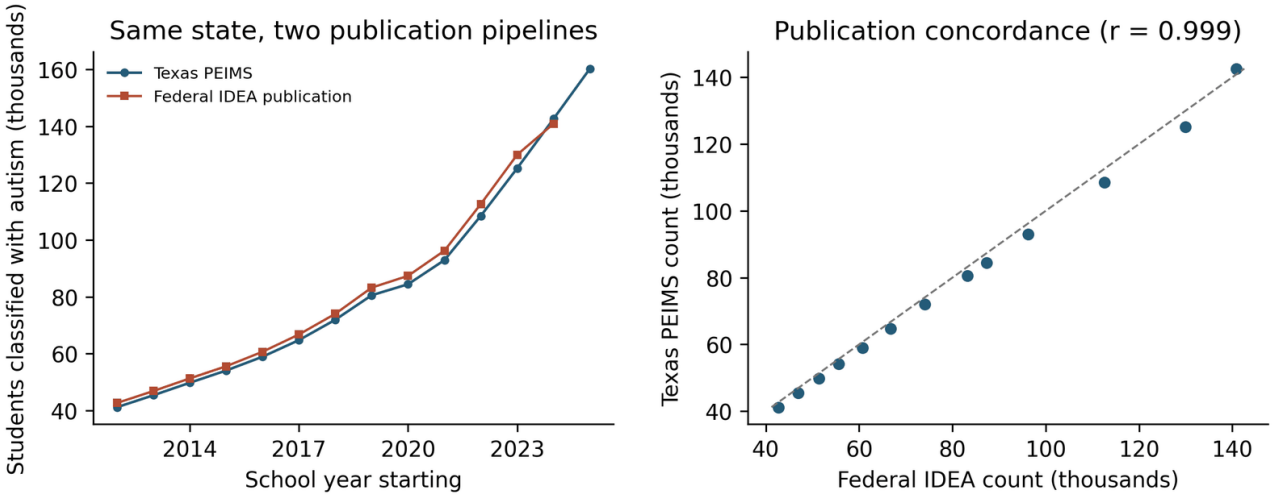


Figure S13. Texas state and federal publication-pipeline concordance

The German RKI/GBE school-entry vaccination-card panel establishes large realized first stages for pneumococcal, rotavirus, and two-dose varicella vaccination. A separate statutory-insurance claims comparison correlates 0.753 across nine regional rankings (Spearman rho 0.783), but it uses a later birth cohort. The comparison validates regional ordering only. No effect estimate is produced without a cohort-aligned ASD outcome (files 277-279).

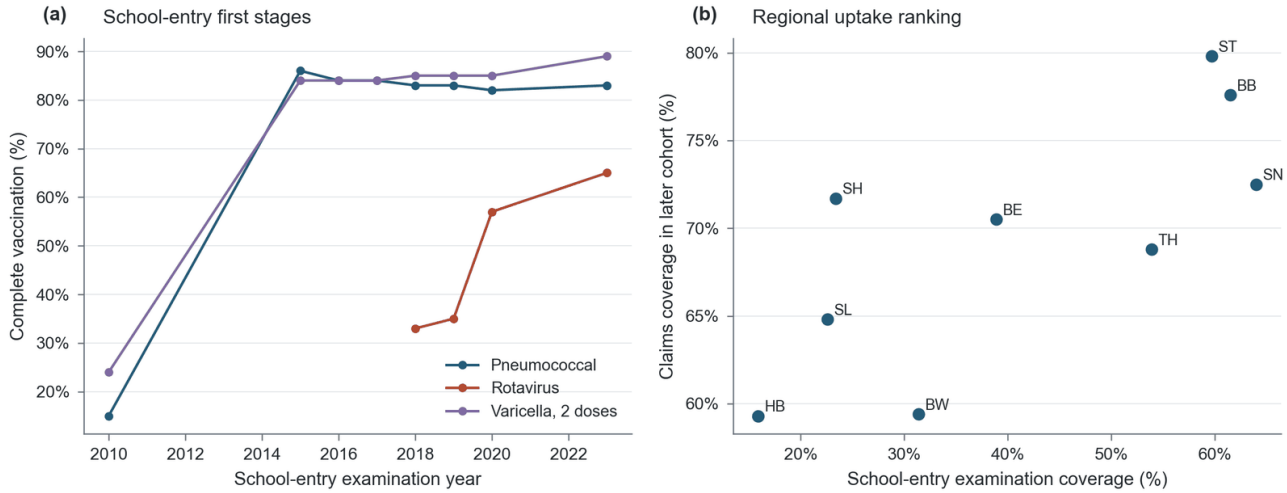


Figure S14. German school-entry vaccination first stages and regional ranking cross-check

S16. Instrument rejection and interaction-event gatekeeping

The proposed PCV7 supply-shock shift-share instrument was tested before any outcome model. Baseline state uptake interacted with the 2004 shortage has excluded-instrument F=0.016 for the target four-dose rate after a baseline-specific trend. The one-dose negative-control exposure moves by 4.44 percentage points with F=16.19. The target strength and specificity gates therefore both fail, and the analysis stops before an outcome IV (files 266-268).

SSI's Danish programme history identifies two temporary hepatitis-B-containing hexavalent additions to an existing DTaP-IPV-Hib schedule in 2014-2017. This on-off structure is a high-value incremental-schedule and reversal event. The public autism report combines ages zero to eighteen by diagnosis year, however, and cannot align exposure to birth cohorts. The event is retained for a Danish Vaccination Register plus patient/psychiatric-register application rather than forced into an age-period-cohort-confounded public regression (files 280-281).

S17. Power and data-access ledger

The reconciled estimate catalog contains 119 specifications. Seventeen have a numerical MDE80 at or below RR 1.10, but only two are both diagnostically eligible and powered for an informative null at their registered reduced-form units: the U.S. PCV7 intensity model (MDE80 1.046) and the autism-specific IDEA diagnostic-substitution DDD (MDE80 1.065). The other fifteen are low-tier outcomes or fail a binding diagnostic. File 265 lists every estimate, MDE, power flag, diagnostic veto, and interpretation so numerical precision cannot be mistaken for evidentiary adequacy.

The dated source-discovery ledger classifies 288 routes: 163 acquired or analyzed, 91 requiring an application, license, purchase, or collaborator, 31 public gaps or incompatible summaries, and three blocked or user-assisted. The canonical source catalog is a separate record layer containing 400 source records. Sections S18-S21 describe the state, school, district, county, and international sources used in this paper. Files 263-264 document the public-data routes, and file 284 ranks the controlled-data opportunities, led by Denmark DDV linkage, U.S. IIS-claims linkage, OpenSAFELY/CPRD/ECHILD, and German cohort-aligned outcomes.

S18. International public-data results

The international inventory contains forty-two country or platform rows, including analyzed panels for Brazil, Peru, Hong Kong, and Taiwan. File 902 separately records Mexico, India, South Africa, other plausible European and Latin American systems, and the MICS/DHS controlled-survey frontier. A source is counted as an estimable design only when exposure and outcome share geography, time, outcome definition, and a defensible cohort or age mapping. This rule excludes regressions based on Mexico's one broad mental-condition cross-section, India's two exposure waves plus one broad-disability wave, and South Africa's nine-province intermittent special-school snapshots. Hong Kong's exact grade-by-year tables permit a modal-age cohort reconstruction, but the national exposure supplies no untreated geography; it is fitted only as an APC-restricted diagnostic. Taiwan's exact county tables permit a broad claim-outcome panel, but only across three high-coverage exposure waves and without a public local autism field.

Brazil's PNI archive contains 1,946,796 municipality-year-measure rows for twenty-six vaccine measures and 1994-2022. The IBGE Census covers 5,570 municipalities, and its join to PNI coverage contains 895,270 rows across 5,554 of them. After excluding one-age-band municipality singletons and using an absorbed estimator that reproduces the balanced state-adjusted first-difference identity, the aligned autism RR is 0.99995 per ten MMR dose-one coverage points (95 percent CI 0.99164-1.00833); younger/future and older nonoverlapping cohort windows are also near one. The conditional 80 percent MDE is 1.0120. MMR, primary polio, and rotavirus estimates against autism and any disability all cross one after family FDR, and pairwise schedule interactions do not survive FDR.

Administrative-denominator sensitivity is reported in full. Capping coverage at 100 and restricting window means to at most 100, 110, or 120 produce intervals spanning one. Requiring every constituent annual value to be at most 100 retains only 84 municipalities with repeated age bands and gives RR 1.276 (1.023-1.592; nominal p=.038). The severe change in sample composition makes this a sensitivity warning, not a preferred correction. Models with singleton fixed-effect cells are excluded, and file 882 verifies the estimator identity; under the valid specification, no rotavirus association is present. Because the outcome is one Census cross-section and age is available only in five-year bands, this remains an exploratory ecological cohort approximation, not a target trial.

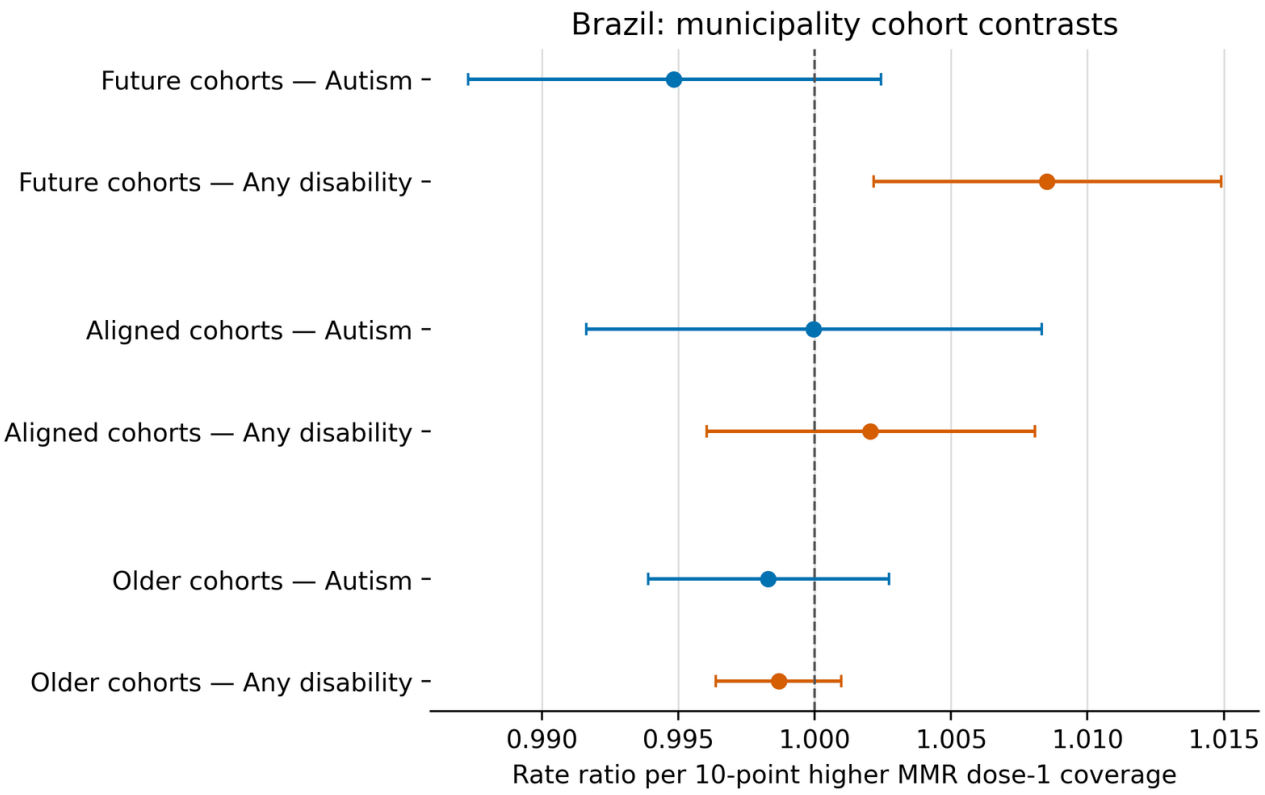


Figure S15. Brazil municipality cohort-window autism and disability estimates

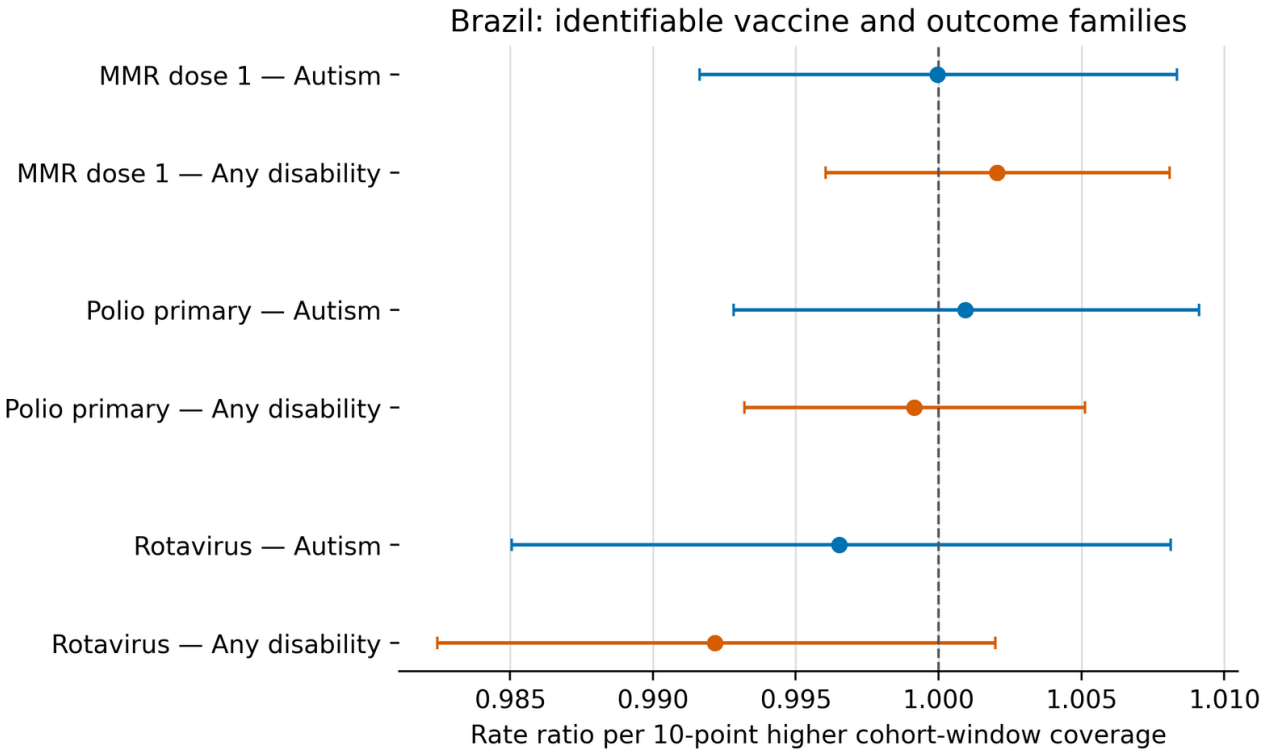


Figure S16. Brazil identifiable vaccine and claim-relevant outcome families

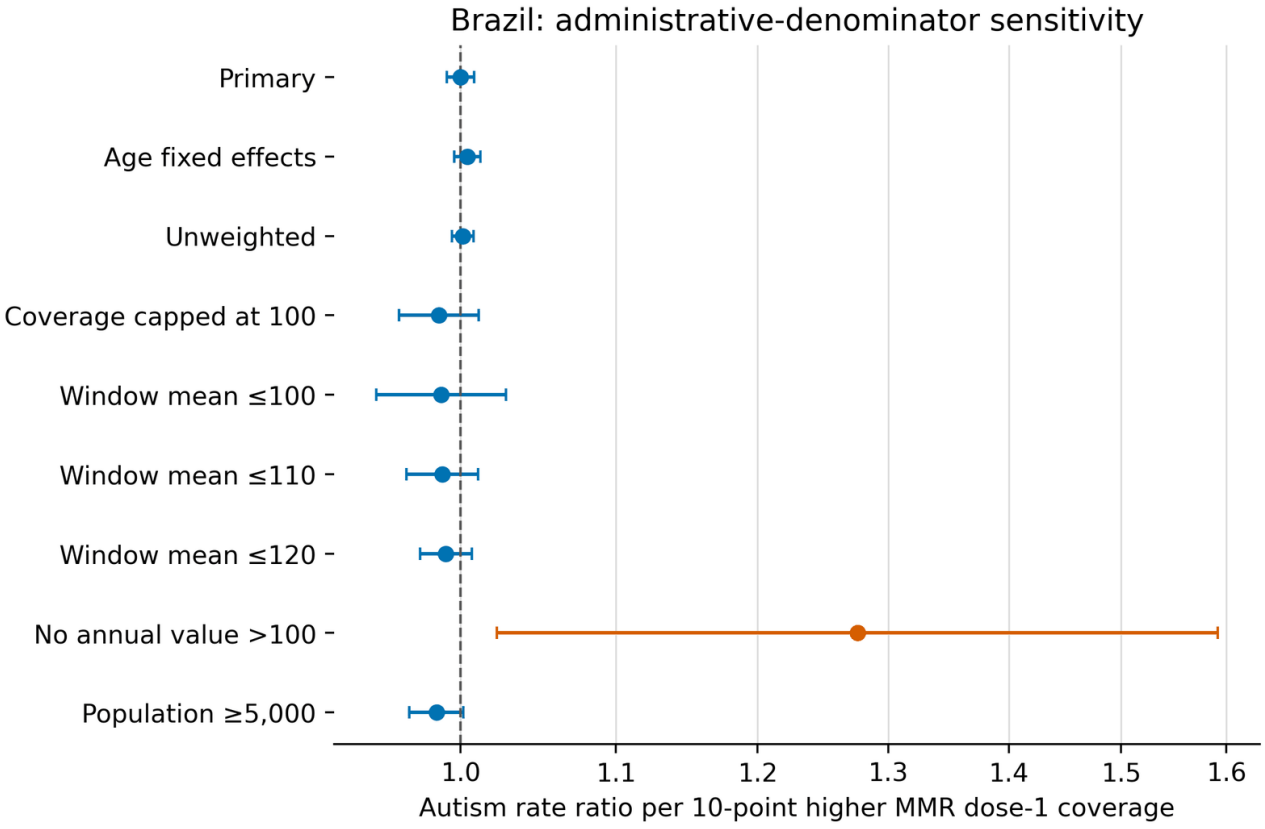


Figure S17. Brazil administrative-denominator sensitivity

The SIH/SUS extension treats epilepsy, viral encephalitis, cerebral palsy, hearing loss, hemorrhagic disorders, broad nervous-system and mental/behavioral diagnoses, and two bowel-obstruction age groups as claim-relevant outcomes, not generic negative controls. Bacterial meningitis and meningococcal infection are positive-control disease families. All

eleven primary contemporaneous estimates cross one and none survives family FDR. The failure of the positive-control outcomes to show protection is reported rather than hidden; it limits validation of the hospitalization design even while the claimed-harm families are null.

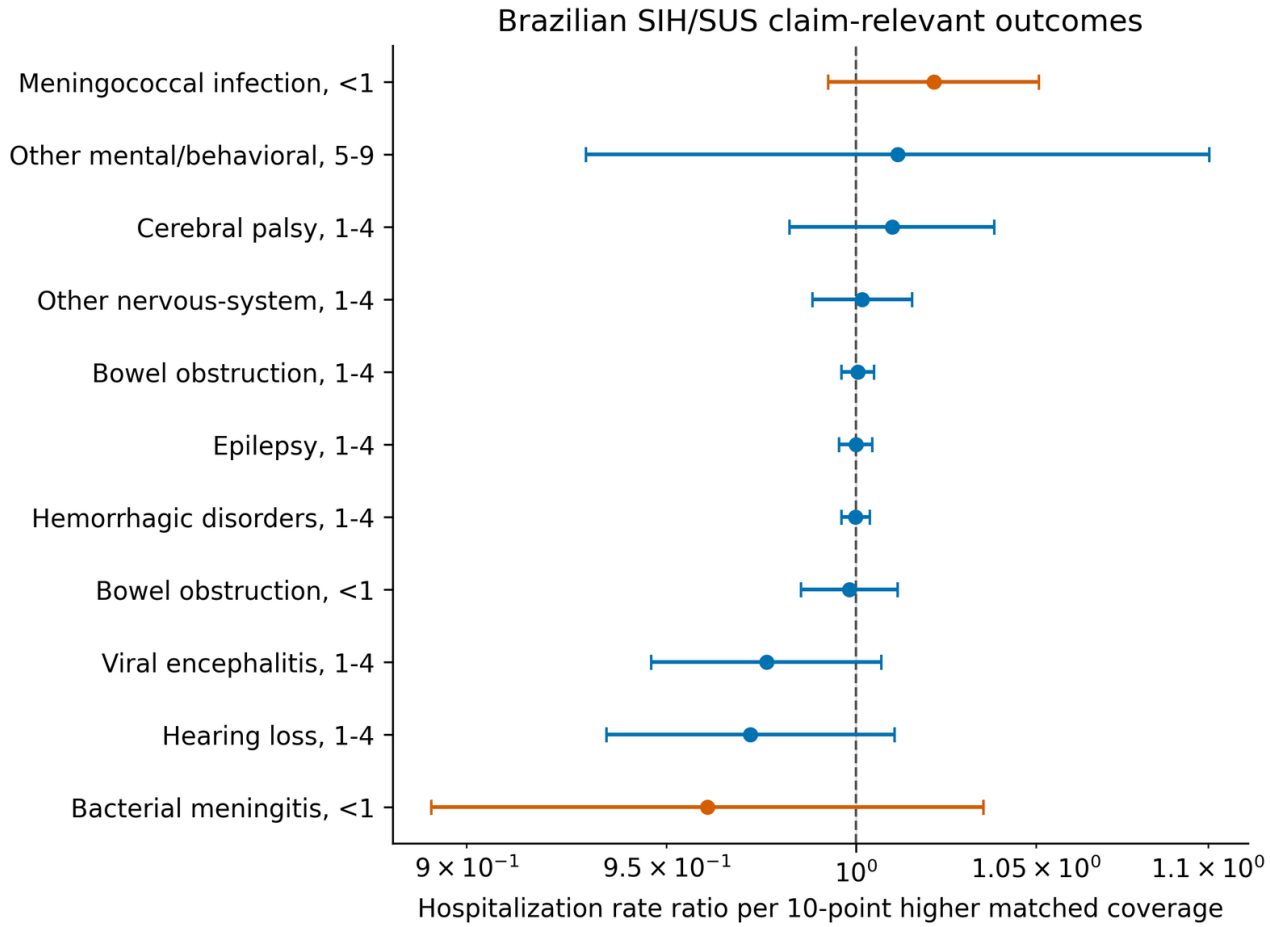


Figure S18. Brazil claim-relevant hospitalization families and positive controls

Peru's positive associations are retained as a failed-diagnostic multiverse. For ages three to five at a four-year lag, autism registration has RR 1.759, hearing 1.769, and vision 1.560 per ten points of MMR's share of EsSalud immunization services. Autism is also positive for future exposure (RR 1.400) and same-year exposure. Exposure is clinic-located service composition rather than population coverage, while CONADIS is a voluntary registration system. The future association, common movement across named outcomes, selective insurance population, and registration timing disqualify a causal interpretation. None of these outcomes is called a negative control; future exposure is the temporal placebo.

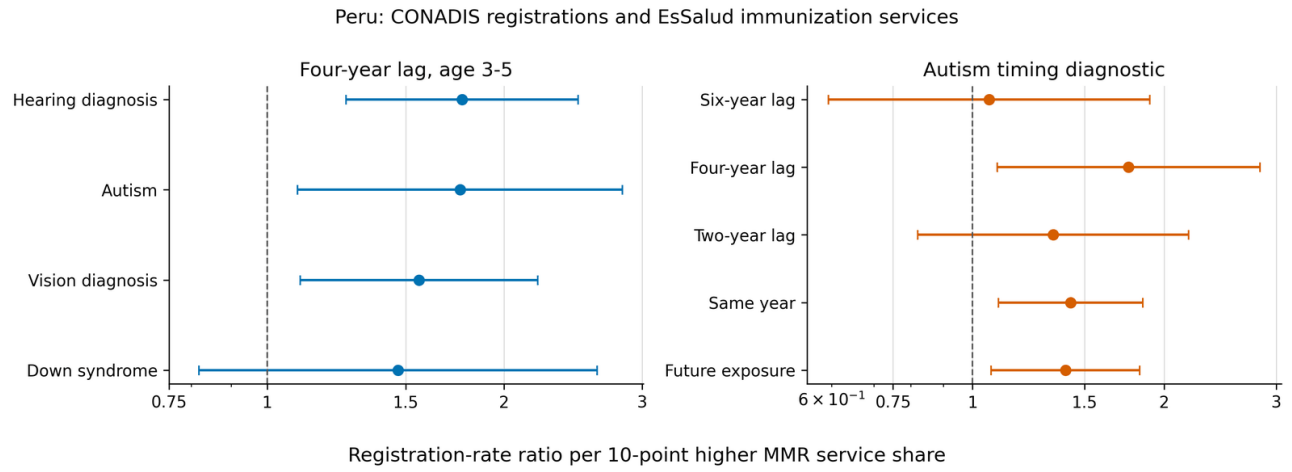


Figure S19. Peru named disability registrations and exposure-timing diagnostics

Hong Kong contributes 504 exact grade-year named-SEN cells, sixty exact grade-enrollment cells, and sixteen modal-age birth-cohort proxies from P1-S6 in 2015/16-2019/20. Official PCV total coverage is 94.8-96.4 percent for the 2009-2012 birth cohorts; pre-2009 is encoded as zero routine-program exposure. In the ASD screen, the aligned RR moves from 0.981 without cohort curvature to 0.998 with quadratic and 0.999 with cubic curvature. The quadratic future-exposure placebo is 0.990 (p=.043), and multiple false cohort cutoffs reject while the actual 2009 cutoff does not. The quadratic conditional MDE is RR 1.023 per ten points, which demonstrates numerical precision but cannot repair APC aliasing, modal-age assignment, school-scope mismatch, or private/catch-up vaccination.

All nine named SEN categories are treated as claim-relevant. Quadratic screens give BH-adjusted discoveries for intellectual disability (RR 1.033; q=.019), physical disability (1.049; q=.016), and the two-year mental-illness series in the opposite direction (0.937; q=.008); ASD, ADHD, hearing, learning, speech/language, and visual impairment cross one. The opposite-signed, non-specific family pattern is reported as evidence of cohort/classification structure under the APC restriction rather than converted into causal vaccine effects.

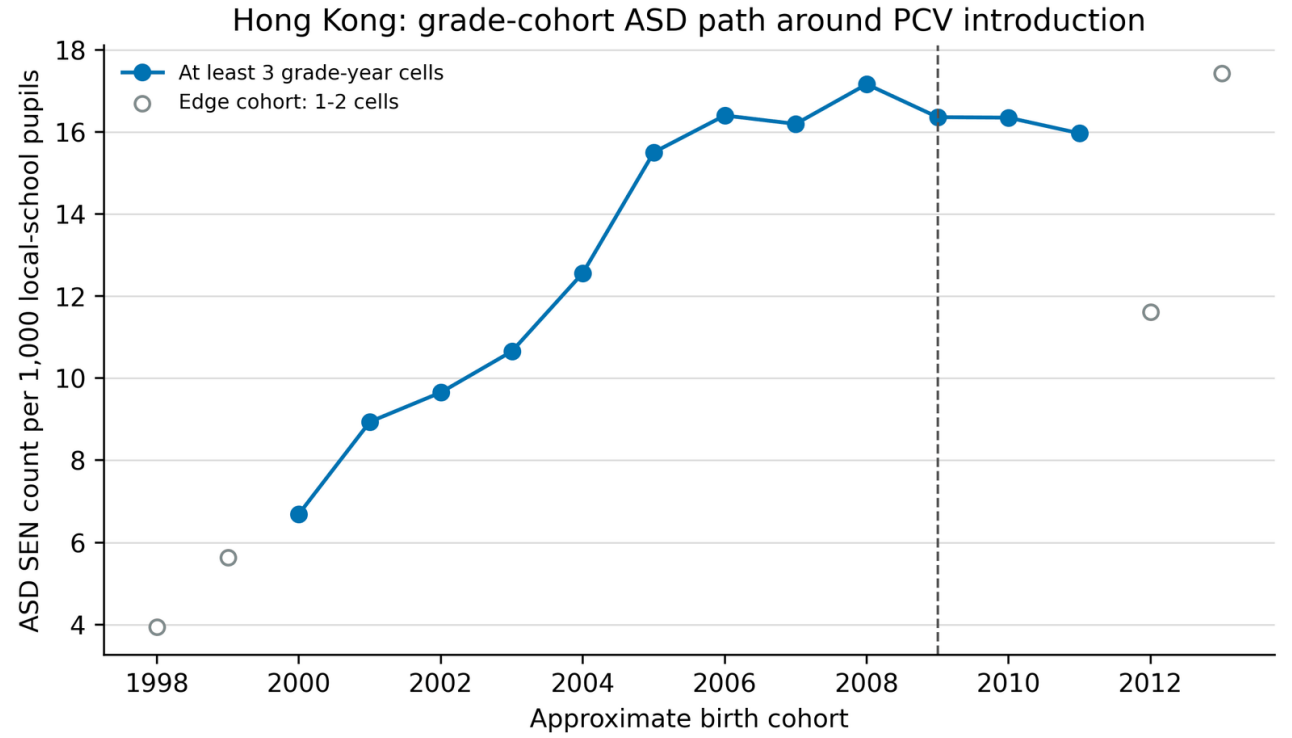


Figure S20. Hong Kong modal-age cohort ASD path around PCV introduction

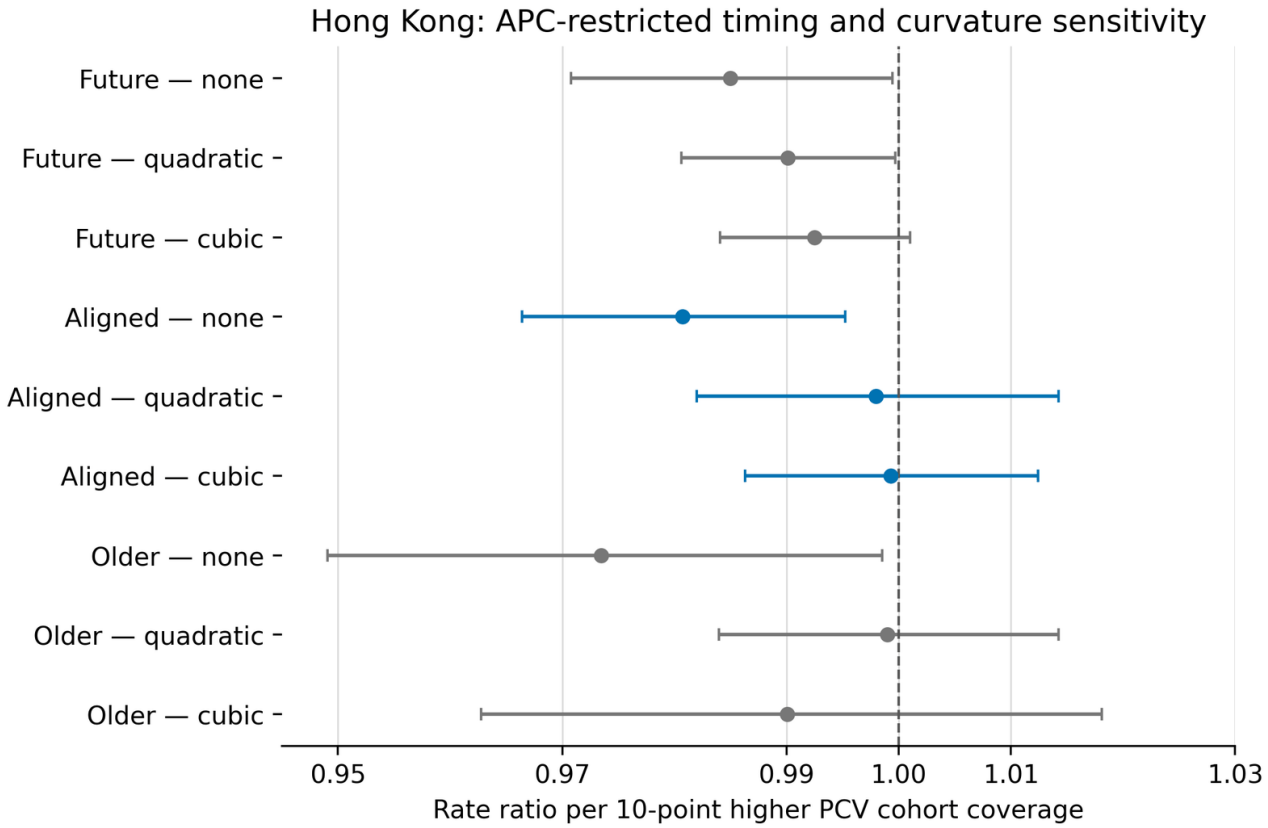


Figure S21. Hong Kong APC-restricted timing and curvature sensitivity

Taiwan contributes a complete low-tier county panel rather than an autism event study. The official Government Open Data releases contain 990 exact source cells: twenty-two counties/cities, three years (2022-2024), and fifteen dose-specific childhood-vaccine completion measures. MOHW workbooks supply six claim-relevant outcomes for 2022-2025: registered disability under age fifteen, developmental-delay reports, current early-intervention cases, children in intervention institutions, children receiving intervention, and intervention-subsidy person-times. Exact household-registration age groups supply the under-fifteen denominators. All source hashes, worksheet mappings, and exact 22-by-year grids appear in files 916-920.

The schedule-mean association with registered disability is RR 0.957 (0.775-1.182) contemporaneously, 1.078 (0.896-1.298) at one year, and 0.913 (0.723-1.154) for future exposure. Every one of the 288 dose/schedule-by-outcome-by-timing tests misses BH FDR, as do all thirty-six centered MMR-pair interactions. Nominal service-count coefficients are sometimes enormous because the within-county schedule SD is only 0.17-0.28 percentage points; the supplement therefore plots one-within-SD effects and reports the unsupported ten-point extrapolation separately. The same-year registered-disability MDE is RR 1.346 per ten points but RR 1.008 over one residual schedule SD. That conditional precision cannot repair the three-wave, no-event, non-autism design. The panel is excluded from causal synthesis, while its broad outcomes remain claim-relevant rather than being relabeled as negative controls.

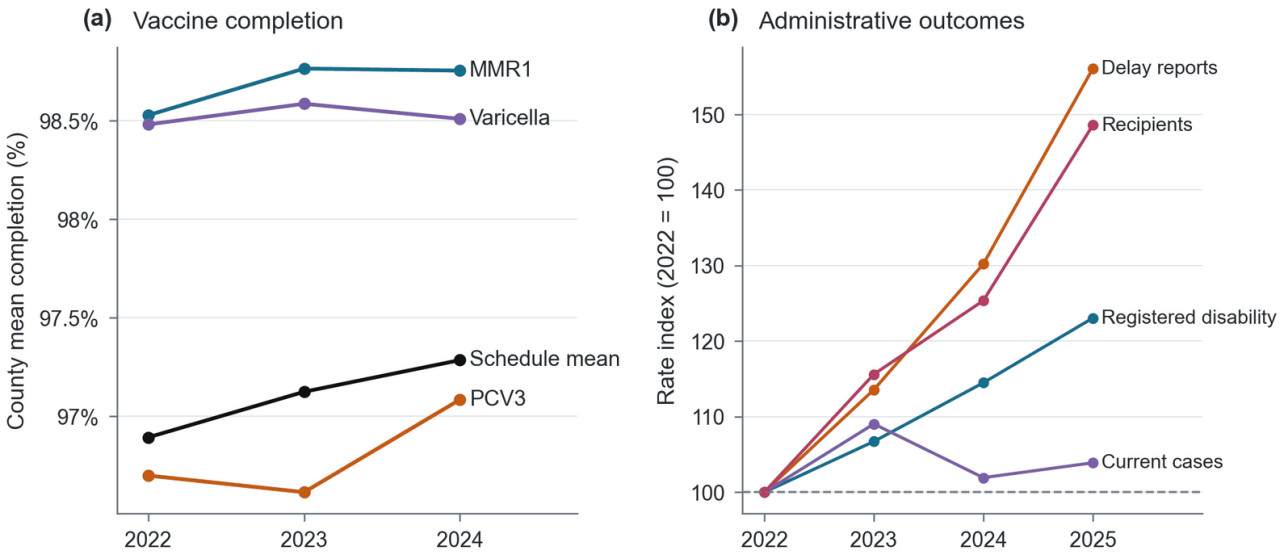


Figure S22. Taiwan county vaccination and claim-relevant administrative outcome trends

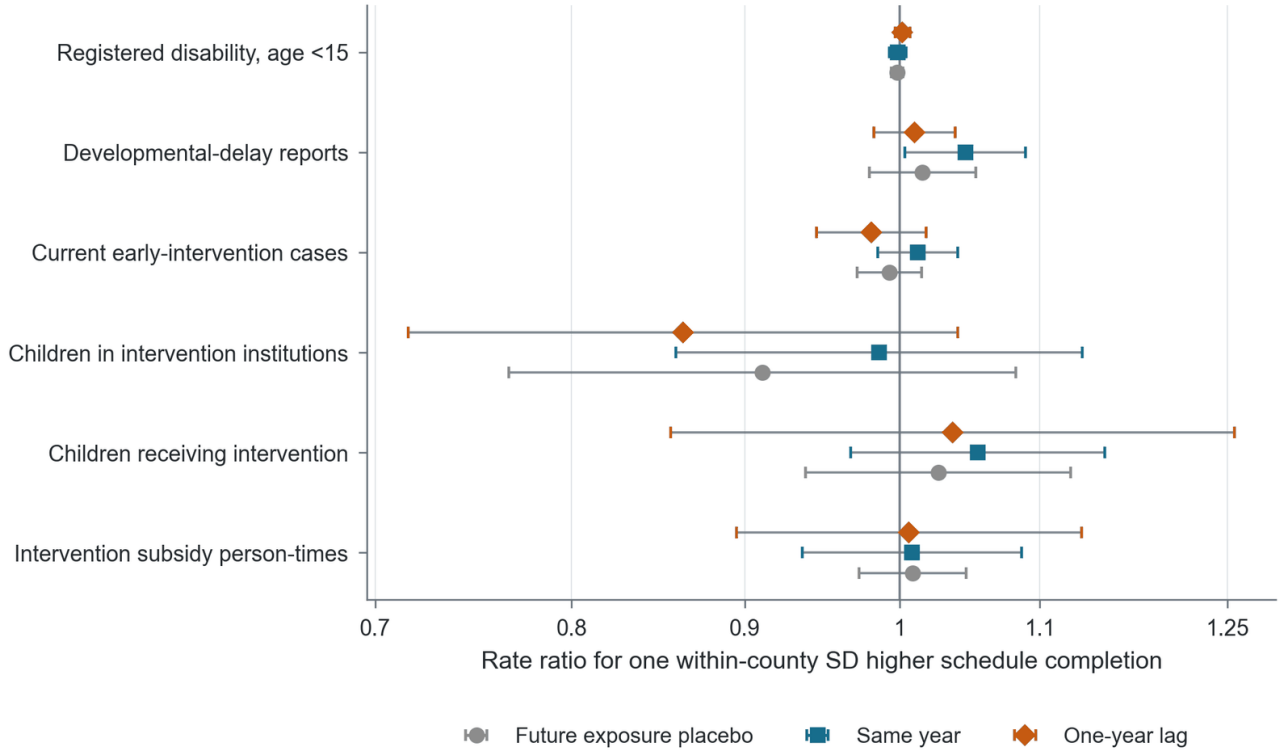
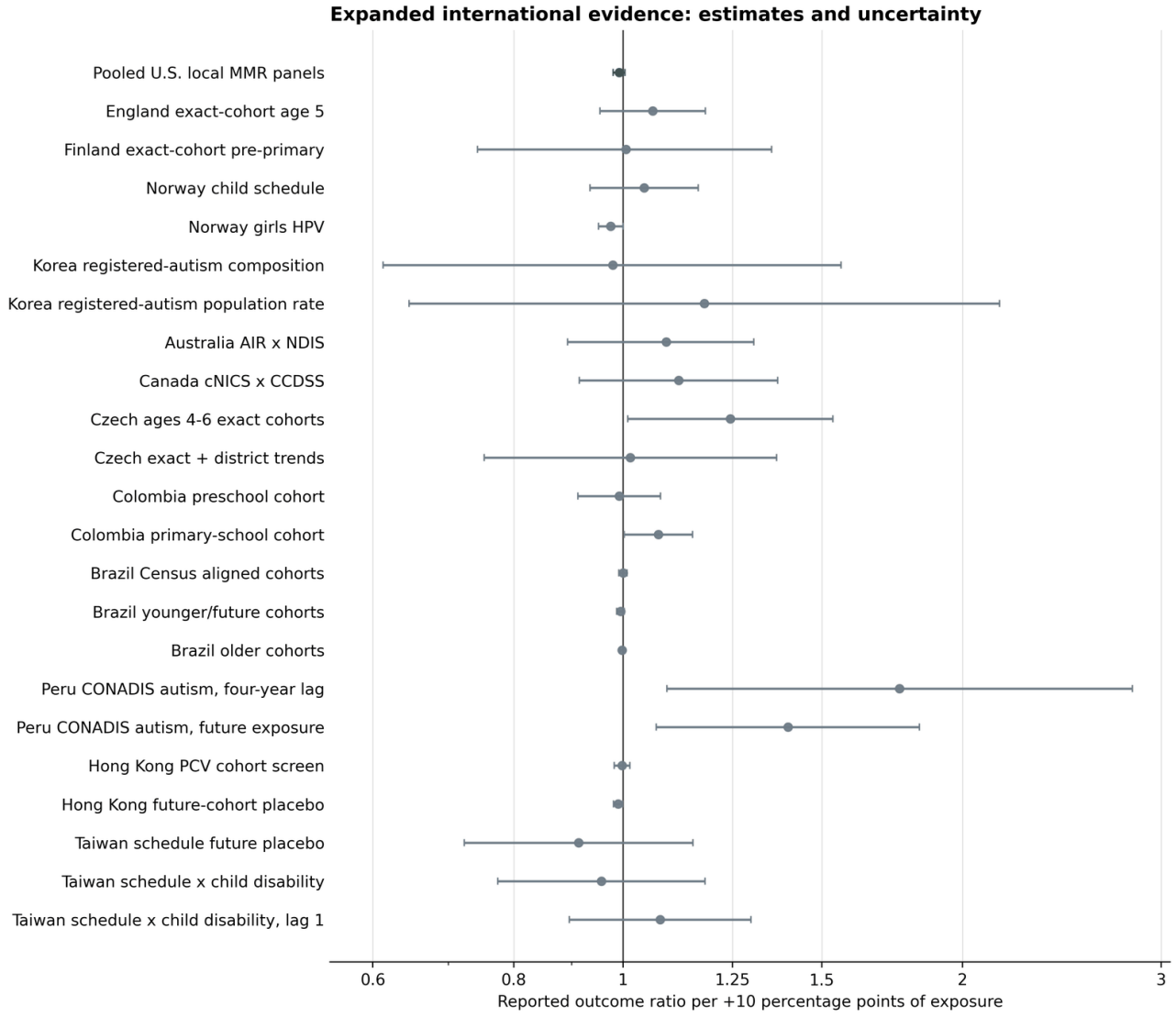


Figure S23. Taiwan schedule-completion associations on the observed within-county scale

The expanded international forest keeps these failed-gate results visible and colors only the mutually exclusive U.S. pooled result as retained synthesis. It does not average incompatible estimands. The power figure likewise reports conditional MDEs after the design labels; high precision is never used to override a failed timing, specificity, or measurement gate.



Conditional ecological estimates; outcomes and timing differ. Rows are not independent causal tests.
Teal identifies the U.S. pooled row; gray identifies other panels and placebo comparisons.

Figure S24. Expanded international public estimates and identification status

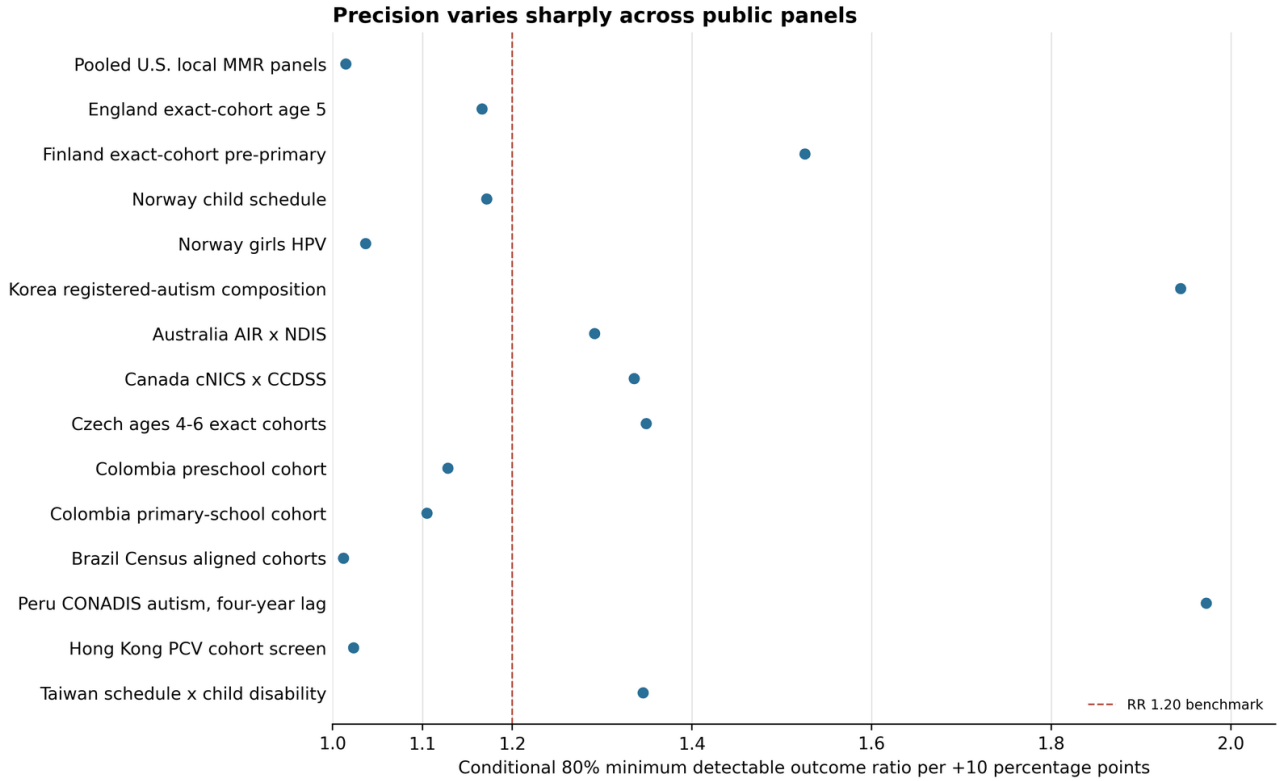


Figure S25. Conditional precision across the expanded public panels

S19. Substate and international source coverage

The data-retention rule is deliberately permissive: an official or reproducible public dataset is preserved when it has descriptive, first-stage, falsification, provenance, or future-linkage utility even if it cannot support the target causal estimand. Entry into a pooled causal result is stricter. Files 1148-1152 separate the source inventory, 2,671 fitted claim-outcome rows, 543 interaction and dynamic-path rows across 43 families, 418 source-tagged power-scenario rows representing 253 underlying power curves, and the pooling gate. Among the 543 interaction rows, 127 have nominal p-values below .05 and 70 survive their sourcewise adjustment; none passes the joint exposure-support, timing, trend, placebo, multiplicity, and causal-design gates. The primary MMR pool uses one geographically disjoint, future-placebo-passing estimate from each of nine jurisdictions; some use the same federal IDEA reporting architecture and are therefore subjected to shared-spine dependence sensitivity rather than specified fully independent. Georgia, Kentucky, North Dakota, Wyoming, and Connecticut enter only a fourteen-design sensitivity. Redundant state-native and federal outcome publications, overlapping California pipelines, alternate school/county/district aggregations, and derived copies are never counted as independent observations merely to narrow an interval. Adjusted discoveries remain visible: the California lagged pair coefficients coincide with rejected future-pair placebos; New York and Colombia discoveries are negative; Peru's positive residual-service interaction inherits the source's failed future/current timing pattern; and Rhode Island measures selected catch-up services. None qualifies as a robust positive injury interaction.

Arkansas files 1089-1105 preserve 13,836 school-years and 3,390 district-years of official exemption reporting, all autism and claim-outcome timings, masking sensitivities, power, and the exemption-timing figure. No family-adjusted claim result is positive. Virginia files 1106-1121 preserve 156,392 school, 27,024 county, and 89,184 ZIP records from ImmunizeVA's derivative of the retired VDH School Immunization Survey. The deletion of the original official files lowers provenance but does not justify discarding the derivative; the county complete-schedule/autism estimates and all four pair interactions are null after correction.

North Carolina has two noninterchangeable products. Files 1122-1131 document the official Tableau filter universe, raw filtered-CSV harvest, school/county/district aggregation, source-grid reconciliation, IDEA joins, all claimed outcomes, lags, future exposure, and power. The raw layer retains 9,583 school-years and 600 county-years; 9,581 school and 596 county vectors are observed, five are explicitly not displayed, and none remains failed. Thirteen official

vectors are nonunit and are flagged without renormalization. The district up-to-date/autism RRs are 0.997 for future exposure, 1.004 contemporaneously, and 1.006 at one year per ten points; all intervals cross one, and none of 192 claim models survives FDR. Files 1140-1147 preserve the ACCIDDA model output embedded by NC DHHS: 1,875 schools and 11,250 school-grade rows, but a modeled rather than raw MMR surface. Its size-adjusted autism cross-section is RR 0.940 (0.814-1.085) per ten modeled points, MDE80 1.228. It is descriptive and is never substituted for the official schedule-status panel.

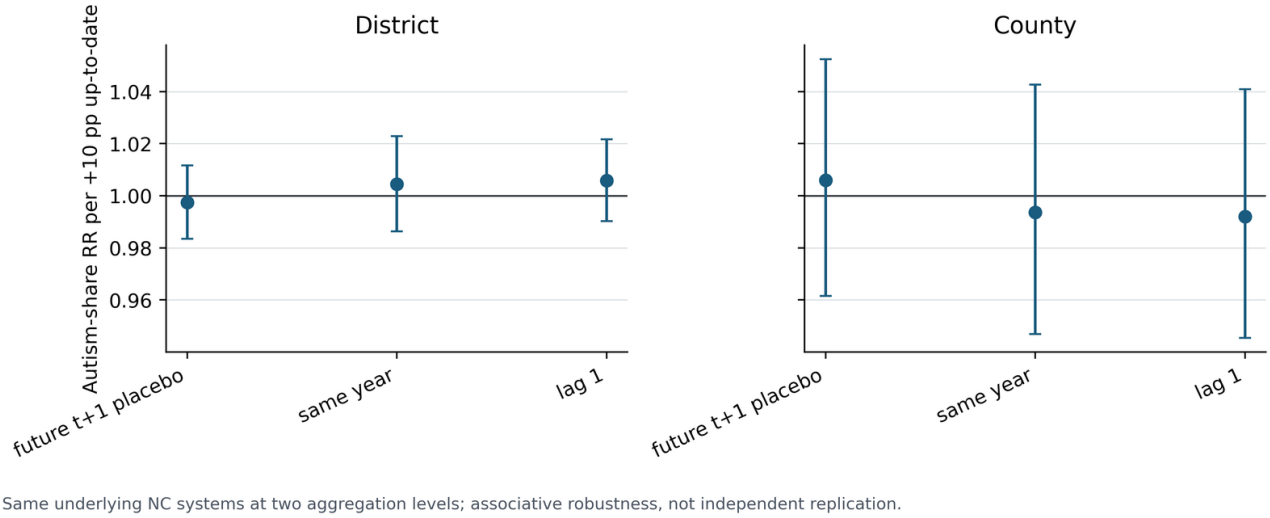


Figure S26. North Carolina schedule status and IDEA autism-composition timing

Wyoming files 1153-1163 archive hashes and parsing decisions for every one of 115 official 2019-2023 county report cards. The normalized exposure file contains 2,460 rows, including a complete 23-county-by-five-year toddler-MMR grid and separately defined school-age and adolescent series. Files 1157-1162 contain 279 claim-outcome models, six pair interactions, four timing-specific power calculations, and a no-failure audit. Autism-composition RRs per ten MMR points are 1.055 contemporaneously, 1.053 at one year, 1.001 at two years, and 1.033 for future exposure; none survives family correction.

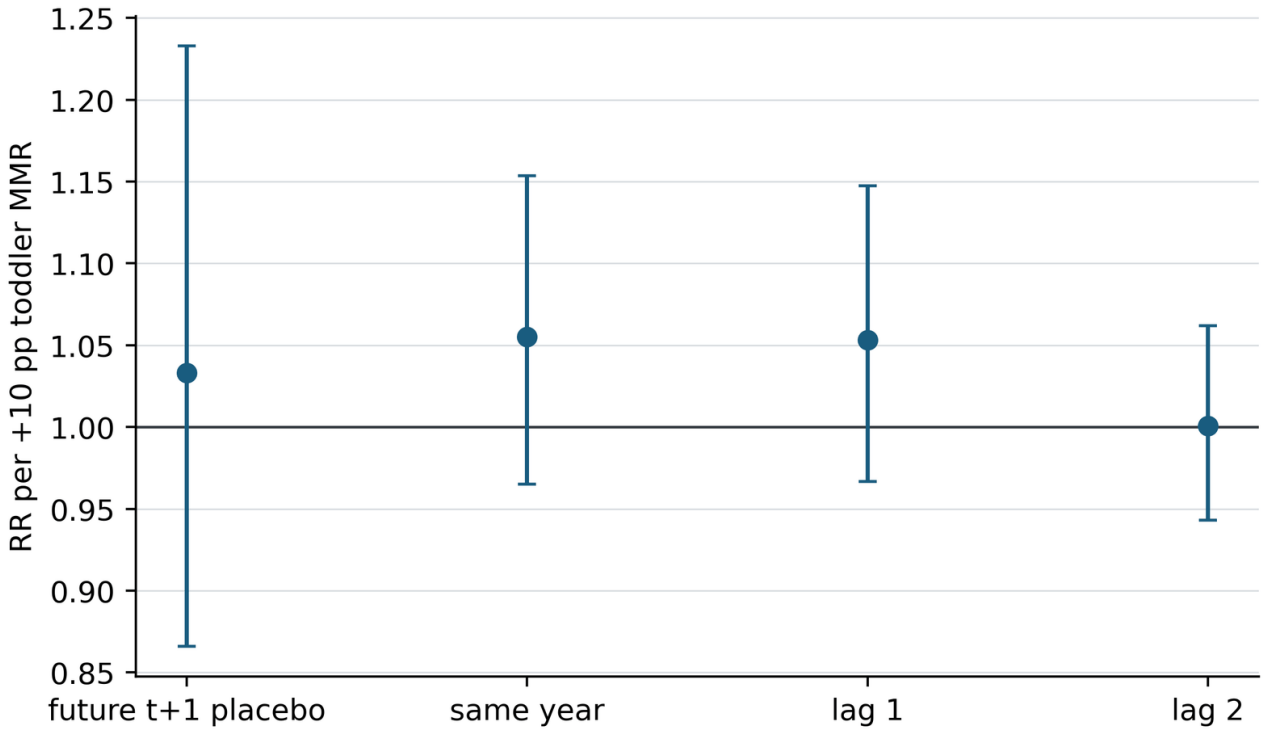


Figure S27. Wyoming toddler-MMR and IDEA autism-composition timing

Rhode Island files 1166-1174 archive and parse all 311 pages of the 2017/18-2019/20 Vaccinate Before You Graduate reports. The 3,421 vaccine-specific rows reconcile exactly to each page total. Hand-audited aliases and conservative fuzzy matching connect 202 of 308 nonstate pages, 98 public schools, and 68.5 percent of delivered doses to 40 districts. Because the numerator is doses delivered to pupils needing catch-up, it measures service use and selection rather than vaccine coverage. The 240 claim screens and twelve service-pair interactions are retained in full. Seven claim tests and two negative pair terms survive within-family FDR, but the mixture of diagnoses, directions, and catch-up selection precludes a coherent causal injury or protection interpretation.

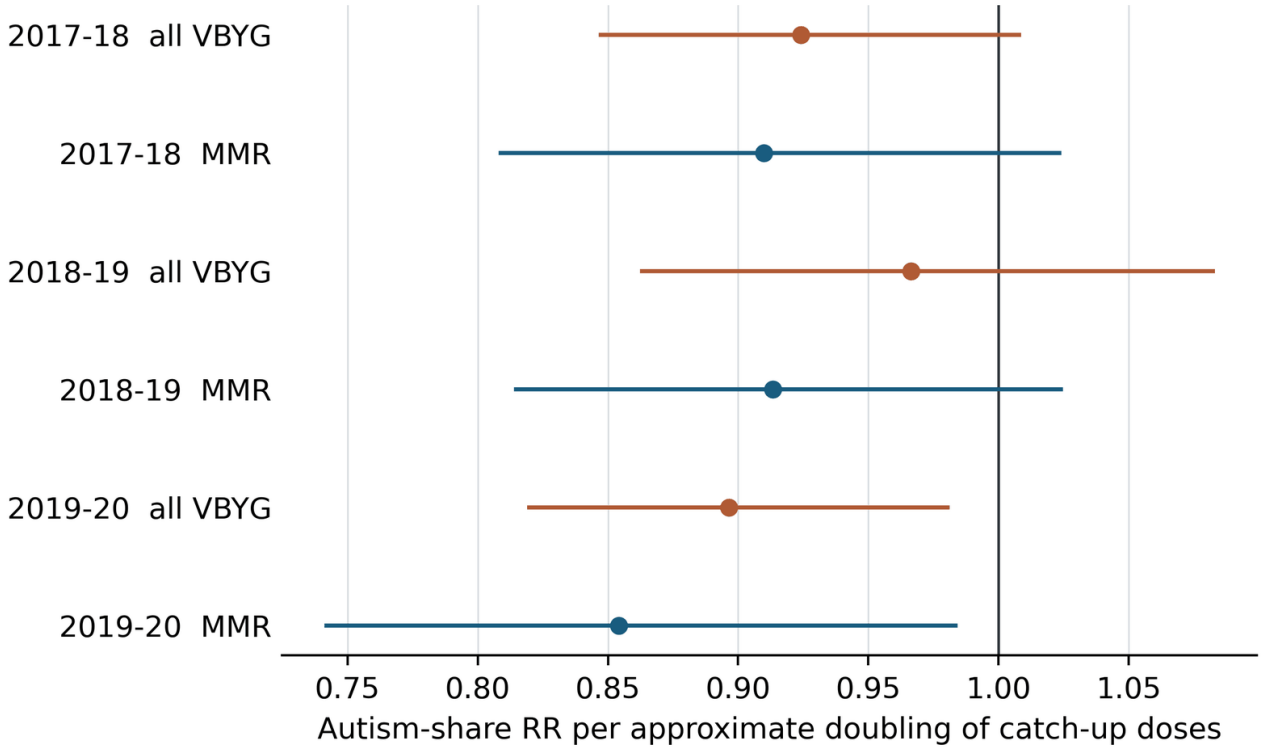
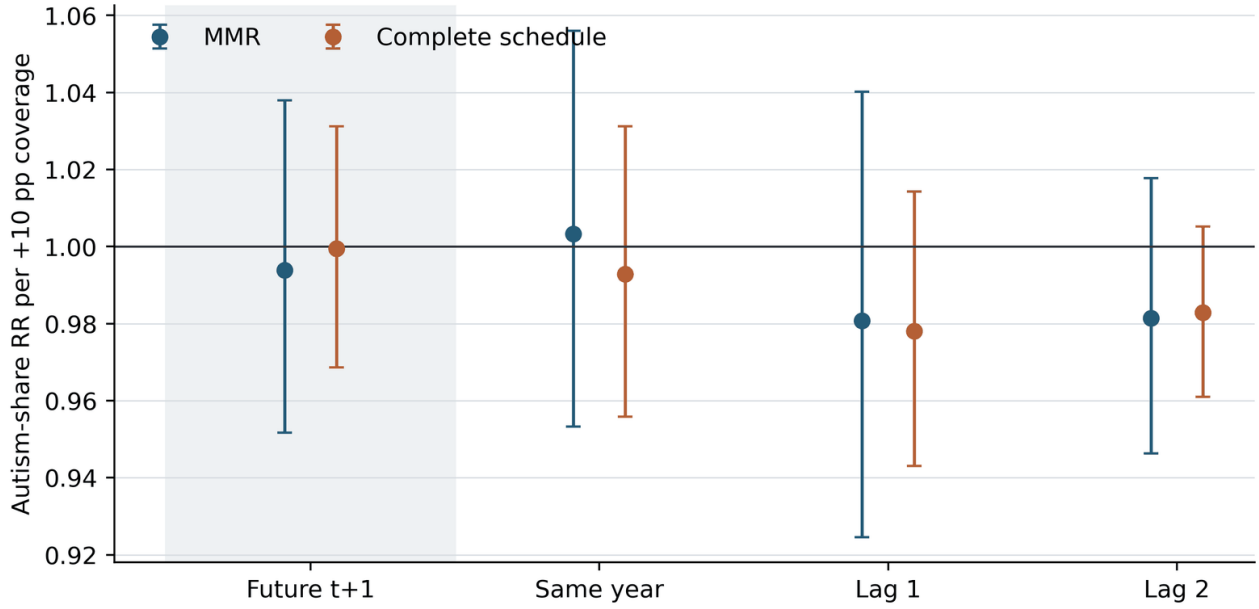


Figure S28. Rhode Island school catch-up services and later autism composition

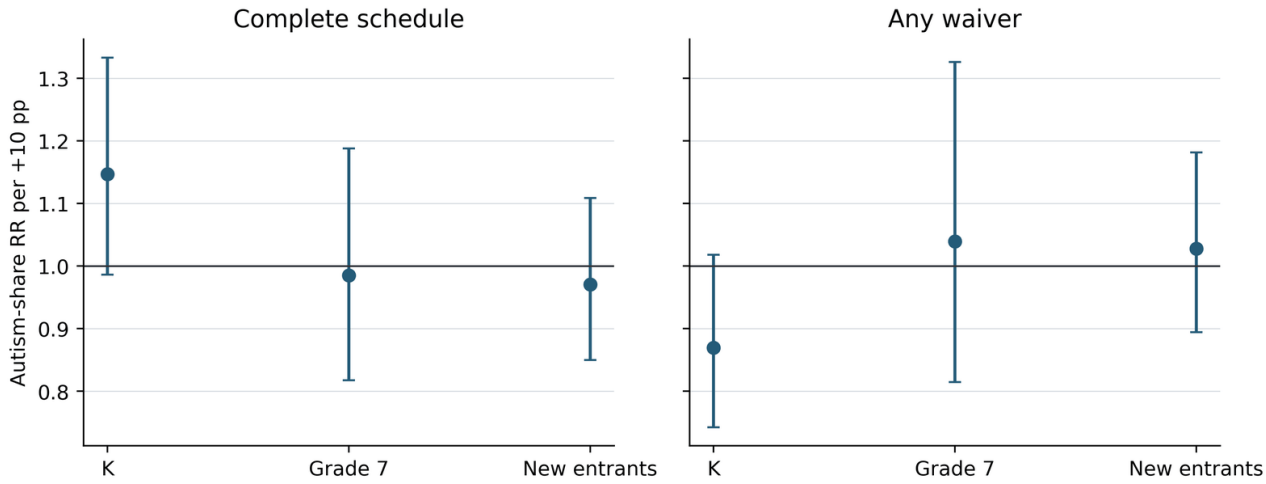
Connecticut files 1189-1197 archive nine official 2017/18-2025/26 school releases, parse 6,140 school rows, and match them to NCES directories to form 1,491 district-years across 179 LEAs. The historical files omit denominators; all district exposures are therefore reported-school means and suppression remains missing. The 224 claim models and ten MMR-pair terms are FDR-null. Autism-composition RRs per ten MMR points are 1.003 contemporaneously, 0.981 at one year, 0.981 at two years, and 0.994 for future exposure; the same-year MDE80 is 1.076.



Matched public-school means; IDEA-enrollment weighted. Associative coverage-environment panel.

Figure S29. Connecticut school MMR coverage and IDEA autism-composition timing

Michigan files 1198-1203 retain all four official 2024 building workbooks: 10,342 building rows and 2,272 matched district-report observations. With only one exposure year, county-adjusted IDEA linkages are descriptive cross-sections, not event studies. Kindergarten complete-schedule coverage gives autism RR 1.147 (0.986-1.333; FDR q=0.568; MDE80 1.239) per ten points; seventh-grade and new-entrant estimates are 0.985 and 0.971. The source releases only combined schedule status and waiver fields, so antigen-pair interactions are not identified.



Michigan 2024 cross-section; county fixed effects. Descriptive, not a timing design.

Figure S30. Michigan 2024 schedule-status and autism-composition cross-sections

Uruguay files 1095-1099 preserve the official 2024 PAI annex: 437 rows, nineteen departments, and twenty-one childhood-schedule or HPV measures. Thirty-one official values exceed 100 percent and remain flagged raw. The single cross-section has first-stage and future-linkage value but no repeated department-cohort autism outcome. Slovakia files 1175-1180 preserve 35 exact Košice-region MMR cells explicitly below 95 percent, ten district-year nonvaccination-reason rows, ten district-year pediatric-circuit coverage ranges, and four audited schedule or campaign events. Unlisted cells are left missing rather than inferred below or above a numeric value beyond the report's threshold logic. Public Slovak outcomes do not supply a compatible district-by-birth-cohort named-autism series, so no eight-region broad-service regression is dressed up as a causal test.

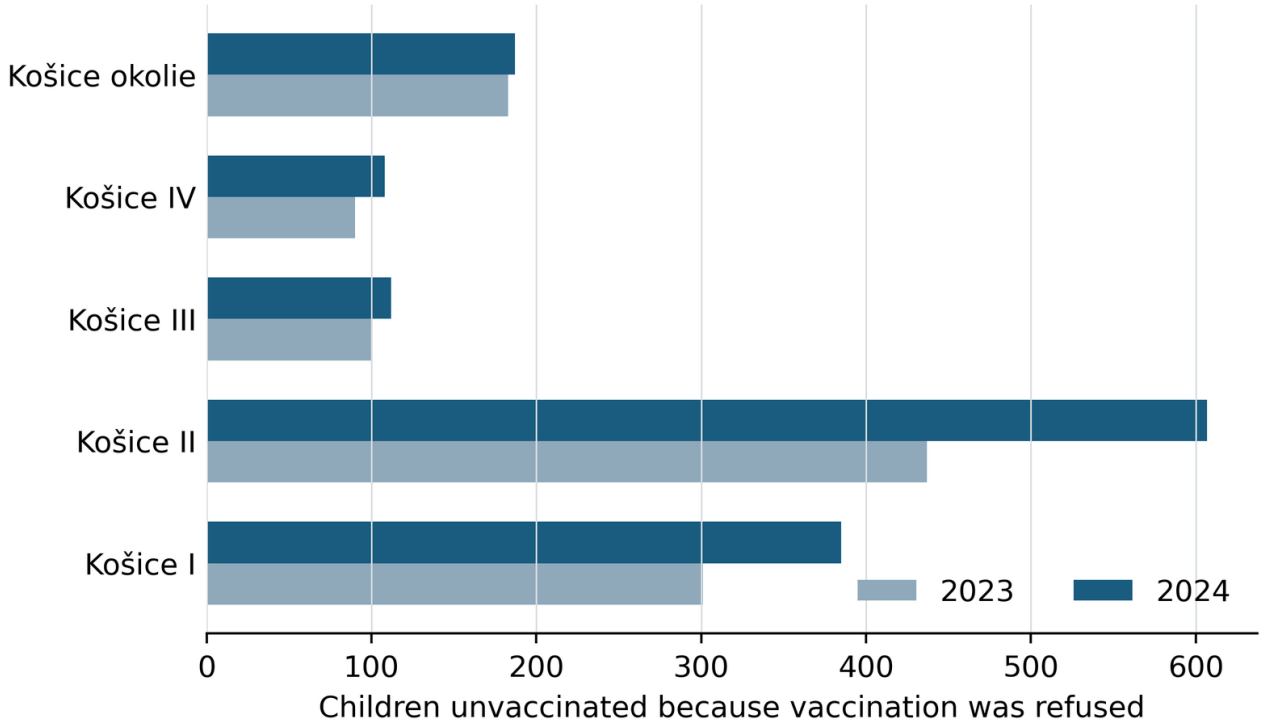


Figure S31. Slovakia Košice-region MMR refusal counts

The fourteen-design aggregate summary in files 1079-1080 gives a contemporaneous fixed-effect ratio of 0.994 (0.984-1.003), with a conditional model-scale MDE80 of 1.013. It is a diagnostic of heterogeneous local gradients, not a common MMR population effect: both outcome denominators and exposure definitions differ. Future and lagged coefficients do not share a validated population-risk interpretation, and the Illinois and Vermont lagged analyses require outcome-year denominators (S67.4). The denominator- and exposure-specific syntheses in S67 replace this mixed-target summary for substantive interpretation.

The worldwide inventory distinguishes more than forty high-information national systems. Detailed official URLs, years, smallest available units, fields, HTTP behavior, archive paths, and binding constraints appear in files 829 and 902. Public pairs are analyzed when geography, time, outcome definition, and a defensible cohort mapping align; otherwise the usable component and exact limitation are documented. Remaining high-value routes require controlled linkage or custom tabulation: U.S. IIS/VSD/CMS or APCD-education linkage; Denmark, Sweden, Norway, Finland, and Iceland registers; German FDZ claims; England's ECHILD/CPRD/OpenSAFELY; Israel and Taiwan health systems; and comparable national health-education linkages. The documented public-data frontier is dated 15 August 2026; it does not imply that no obscure local file exists in every one of roughly 195 countries.

S20. Regional, municipal, trust, county, and worldwide data

The archive retains data that cannot identify a vaccine-autism effect when they provide exact local exposure fields, broad disability outcomes, source-independent publications, administrative mappings, descriptive worldwide coverage dispersion, or explicit publication-grain constraints. Only compatible, mutually exclusive, timing-valid estimates can alter a primary pool. Scripts 252-265 and files 1293-1368 implement that distinction. The standardized coefficient-and-constraint ledger contains 538 rows, each with its native effect scale, estimate and interval, multiplicity result, conditional MDE where estimable, ecological constraint, individual-translation requirements, binding identification limitation, and a specified zero causal-pool weight.

Flanders. Official education files contribute 54,885 campus/school rows, 7,371 municipality rows, and 505 province rows. The outcome is Type-9 special education, which identifies ASD without intellectual disability among pupils receiving that program and is neither all ASD nor a population diagnosis. Twenty-four five-province change screens produce three nominal p-values and no BH discovery. Young-child MMR coverage is positive both for the contemporaneous change (+9.82 Type-9 pupils per 10,000 for one coverage point, p=.034, q=.275) and for the future

outcome change (+6.28, p=.014, q=.164). Five clusters, program selection, and the future movement preclude a causal interpretation.

Lithuania. The claims panel contains 90,327 municipality-year diagnosis rows. Public pupil-health files cannot be joined because the masked identifiers are table-specific; that is recorded as a schema failure rather than imputed. Sixty-four claims associations yield six nominal p-values and four BH discoveries. The municipal autism association is RR 1.097 per ten points of any-vaccination intensity contemporaneously (0.986-1.221; p=.090) but 1.134 for future coverage (1.024-1.256; p=.016; q=.203 in the broad family). The stronger impossible-timing association, shared claims/healthcare recording, and lack of a policy event disqualify it from causal synthesis.

Slovenia. Official sources contribute 810 exposure cells across nine health regions and 5,512 municipality-year education-outcome rows. The outcome is participation in adapted programs, not named clinical autism. Forty main-effect and five vaccine-pair estimates yield eight nominal and two family-adjusted main-effect discoveries; all five interactions are null. MMR is 1.046 contemporaneously (0.996-1.099) and 1.044 for future exposure (0.973-1.120). Geography mismatch, nine clusters, broad program scope, and no stable exposure event govern interpretation.

Northern Ireland. Five health-and-social-care trusts provide the only lower-level units, and the autism/SEN publication changes from a special-education register to a medical register in 2019/20. The post-break MMR2 association is RR 1.776 at one year (1.332-2.368), while future exposure is also positive at 1.825 (1.207-2.759). Eighty-five main, outcome-family, and interaction rows yield thirty-five nominal and twenty BH discoveries, but the residual exposure SD is about one point and the displayed ten-point effect is extrapolative. Five clusters, a definition break, positive future exposure, and broad SEN components rule out a causal event interpretation.

Croatia. The official county panel contains 2,764 vaccine rows across twenty-one counties in 2016-2024. The available local outcome is broad disability among ages zero to nineteen in 2022/23; F84 autism is available only nationally. Fifteen county screens yield one nominal result and no BH discovery. A single broad-age outcome wave cannot identify cohort timing, so the panel is retained as a first-stage and outcome-grain audit.

Latvia. Official sources contribute 364,428 pediatric-facility aggregate rows, 44,861 school-program rows, and 2,013 municipality-category-year outcome rows. Only the current facility directory is public; address-based linkage reaches 22.5 percent of dose-weighted activity and cannot reconstruct historical catchments without error. Of 260 fixed-effect screens, thirteen are nominal and none survives BH correction; all twenty pair interactions are null. The large row count does not compensate for weak historical geography, broad outcomes, and nonrandom provider catchments.

Romania. INSP's February 2022 and 2023 reports provide county MMR2 coverage at age five and dTpa coverage at age fourteen. INS TEMPO SCL103E supplies 1,258 county-years of pupils in special primary or lower-secondary schooling, 1995-2024. The outcome is a broad special-school share, not autism, and coverage has only two report years. The MMR contemporaneous RR is 0.982 per ten points (0.960-1.003; p=.099); dTpa is 0.997 (0.977-1.017). The MMR prior-outcome placebo is 1.010 (0.980-1.040); subsequent-change RRs are 0.999 for MMR and 1.019 for dTpa. The MMR-by-dTpa term is 0.980 (0.942-1.020). None of seven rows is nominal or family-adjusted significant. Official school-level workbooks are catalogued, but the public data portal timed out during reproducible acquisition; this is an access-status entry, not evidence that the files do not exist.

Turkey. Ministry of Health Statistics Yearbooks provide DTP/IPV/Hib3, HBV3, and MMR coverage for twelve NUTS-1 regions, 2017-2024. The 2017 PDF's embedded chart font is broken, so the exact dark-blue chart labels were transcribed from rendered official pages 105-107 and the method is flagged in the source audit. Outcome yearbooks provide first child special-needs report incidence at at least forty-percent severity, residence-based, for 2020-2024. This is broad disability-report incidence, not autism. DTP and HBV future, current, and lag coefficients all cross one. MMR is positive contemporaneously, RR 1.119 (1.012-1.236; p=.028), and at lag one, 1.236 (1.028-1.485; p=.024), but near one at lag two and negative at lag three, 0.893 (0.829-0.963). All three schedule-pair interactions cross one. One of eighteen rows survives BH correction, but the positive-short-lag/negative-long-lag reversal, twelve clusters, five outcome years, broad outcome, and absence of a discrete event make the pattern identification-ineligible.

Multiplicity accounting. The eight panels contribute twenty-seven BH-adjusted main-effect discoveries: Northern Ireland twenty, Lithuania four, Slovenia two, and Turkey one. Flanders, Croatia, Latvia, and Romania contribute none. This count is reported because the total-evidence rule applies equally to positive, negative, and null coefficients. It is not used as a vote count. BH correction controls repeated sampling within a specified family; it cannot correct a future-exposure association, incompatible cohort ages, outcome-definition changes, geography mismatch, or an absent counterfactual. Accordingly, all 538 rows are visible and all carry zero primary-pool weight.

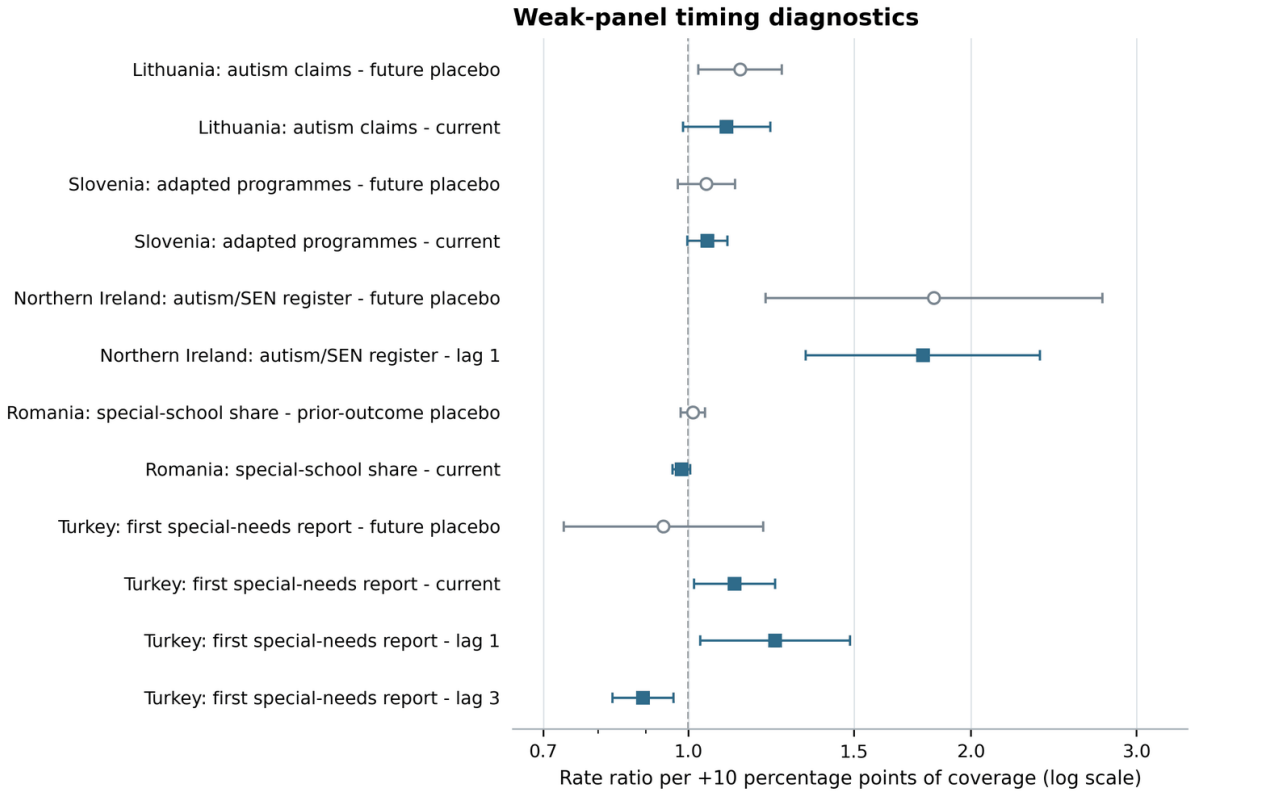


Figure S32. Regional and municipal timing diagnostics

The figure makes the binding timing problem visible by pairing current or lagged coefficients with future-exposure or prior-outcome estimates. Lithuania's future autism association exceeds its current association; Northern Ireland's current and future estimates are both positive; Romania is near one in both directions; and Turkey's positive short lags reverse at lag three. These patterns are inconsistent with a single stable exposure-to-outcome lag, although they do not prove that every product-outcome effect is zero.

WHO worldwide subnational distribution. The official WIISE public endpoint MT_RI_IND_DIST_LONG yields 256,104 nonmissing country-year-indicator records across 235 countries and areas, 1999-2025, and 198 indicators. WHO's public page describes annual coverage data for more than 29,000 subnational units in 147 countries in 2025. The endpoint available without credentials exposes country, year, indicator, and counts or percentages of districts falling in coverage bands; it does not expose the named district identifier needed to join a local outcome. The OData metadata contain 163 entity types and were searched for administrative-level-two fields. The admin2-like entities are outbreak datasets, not routine immunization coverage. This is therefore a completed worldwide descriptive acquisition and a precise publication-grain block on named local causal linkage, not a missing regression.

Coefficient and individual-scale reporting. The publication workbook keeps the model-native numerical ledger and translation scenarios distinct. The nine-design aggregate calculation, 0.992 (modified-HKSJ 0.980-1.004), mixes special-education composition, enrollment-based outcomes and schedule compliance. Its algebraic conversion to 0.925 (0.830-1.038) under an all-child risk-mixture formula does not meet that formula's denominator requirements and is not an individual-risk bound, even conditionally on homogeneous vaccine effects. Autism's share of special education cannot be substituted for autism's population prevalence without the special-education enrollment rate and its covariance with the share. The separated local results and exact denominator identity appear in S67; archived translation files retain provenance but have no individual-risk interpretation.

Source-frontier conclusion. The sources described above do not alter the primary PCV7 result, the historical mixed-target local summary or its model-scale power calculations. Their valid interpretations broaden geographic and outcome coverage, identify false-timing and definition-driven positive patterns that a selective review could miss, and establish exact public-release constraints for future work. The documented public-data frontier is dated 15 August 2026;

it is not a claim of literal global exhaustiveness. Material gains beyond this frontier require custodial access, custom tabulation, historical crosswalks, or linked person-level registries.

S21. School, district, province, territory, and claims data

Purpose and scope. A source is included if it supplies a distinct unit, longer history, exact named outcome, independent official publication, denominator, suppression rule, crosswalk, or precisely reproducible access constraint. An estimate enters the primary causal pool only if its population, geography, cohort timing, counterfactual, and diagnostic behavior are compatible with the specified synthesis. Files 1370-1484 preserve the sources, model rows, constraints, power, and validation. This inventory has fifteen jurisdiction or source-system entries, all with primary causal weight zero.

United States residual frontier. Scripts 267-272 reconcile all fifty states, the District of Columbia, and Puerto Rico, Guam, the Northern Mariana Islands, the US Virgin Islands, and American Samoa. Sixty-eight official assets were attempted; fifty-nine were acquired with raw bytes and SHA-256 hashes, and nine exact links returned HTTP 404. Alaska's state-native DEED district archive supplies thirty-one annual named-autism waves, 1994-2024, totaling 1,685 district-year observations; 1,143 suppressed cells are retained as lower/midpoint/upper intervals, and every published state total falls inside the district bounds. VacTrAK supplies 480 recent region-by-age/series cells, but only four snapshots beginning in 2024 and public-health-region rather than school-district geography. The design gate therefore fails on overlap and geography. West Virginia contributes a 2021/22 statewide autism count of 3,108 and all-disability count of 43,170. Guam contributes autism snapshots of 169 in 2013, 181 in 2015, and 252 in 2019. CNMI contributes 2025 up-to-date coverage of 71.9 percent for ages zero to four and 87.4 percent for school age. These one-unit snapshots are evidence records, not synthetic counterfactuals.

British Columbia. Scripts 273-275 acquire 8,702 BCCDC vaccination rows covering Age 2, Age 7, Grade 6, Grade 9, sex/international-status strata, and 2016-2025; 331,330 DataBC diverse-ability rows for 1996/97-2025/26; 40,768 exact code-G ASD rows; and 206,058 grade/enrollment rows. Historical and current school coordinates joined to official health service delivery area polygons yield 90,336 school-year mappings with 99.96-percent polygon assignment. In the 2017-2025 analytic overlap, 99.0 percent of enrollment rows and 99.3 percent of disability rows map. Lower/midpoint/upper reconstructions assign 1/5/9 to masked cells; these bounds contain 99.62 percent of unmasked native district-category totals overall and 99.78 percent for ASD. The reconstruction and a native-district/partial-HSDA robustness version each produce 180 screens. Representative school-reconstructed midpoint estimates per ten coverage points are -0.127 ASD-rate percentage points for age-two MMR (SE 0.173), -0.118 for age-seven measles (SE 0.106), -0.211 for age-two rotavirus (SE 0.119), and -0.0099 for grade-nine MenACWY (SE 0.0201). Grade-six HPV has only three usable years and is not fit. Age-two refusal has a strong inverse association that is explicitly marked noncausal/confounded. The outcome pools all K-12 grades rather than tracking the milestone cohort, coverage is not assigned to individuals, and no policy shock supplies exogenous timing. All 360 coefficients receive zero causal weight.

Bosnia, Bulgaria, Serbia, North Macedonia, and Montenegro. Federation of Bosnia and Herzegovina bulletins supply 561 canton/entity-year-antigen rows in 2022-2024, with planned target, vaccinated count, and percentage; all 561 arithmetic checks pass. Three waves, a 2024 combined-series label change, and no canton childhood F84 outcome prevent an event study. Bulgaria supplies 224 Census cells for under-sixteen disability degree across twenty-eight districts in 2021 and national immunization reports for 2022-2024 plus 2025H1. The outcome predates the exposure window, is not autism-specific, and exposure has no district variation. Serbia supplies 1,749 municipality-vaccine rows in the complete 2022 Batut annex across 159 municipalities and twenty-five districts, with 2018-2023 national reports archived; located child morbidity files expose only broad ICD chapters. North Macedonia supplies fifty-four coverage and fifty-four AEFI reporting rows. Its MMR1 series falls from 93.3 percent in 2014 to 62.7 percent in 2025, but there is one national unit and no compatible named-autism outcome. AEFI totals of 17, 23, 4, 5, 17, 25, and 33 in 2018-2024 mix IPH Form 3 and later MALMED-only sources; they are spontaneous suspected reports without administered-dose denominators.

Montenegro is a particularly useful future first stage. Eleven official Health Statistical Yearbooks yield 3,330 municipality-measure rows, 2012-2022, from eighty-eight source tables, including BCG, DTaP, polio, hepatitis B, MMR, school-age boosters, and influenza with denominators and vaccinated counts. Eleven of eleven MMR waves are present; no vaccinated count exceeds its denominator. National MMR1 coverage falls from 90.3 percent in 2012 to 64.0 in 2015, 46.6 in 2016, 23.3 in 2020, 18.3 in 2021, and 32.9 in 2022. Public education releases provide only broad national special-educational-needs totals, and catalogue searches locate no stable municipality autism series. DTaP3 and

polio3 coverage are exactly equal in 90.38 percent of matched rows and correlate .999735; first-booster coverage is exactly equal in 78.68 percent and correlates .999225. Their interaction therefore lacks independent support and is not estimated as biological synergy.

Thailand. Scripts 279-283 retain twenty-three exact source/access entries and thirty-four raw-manifest files. The exposure package contains 154 province MMR2 rows (seventy-seven province slots in completed FY2025 and live partial FY2026, with seventy-six numeric each and Bangkok explicitly missing) and 1,272 local rows from Tak, Satun, Kamphaeng Phet, Chachoengsao, Chumphon, NSO MICS, and NHSO sources. The outcome package contains 7,066 exact-count rows, including Sakaeo subdistrict autism in 2017-2023, Nonthaburi school-by-semester autism/control categories in 2022-2024, Nong Khai and Phitsanulok district panels, and further province/local files from Tak, Satun, Pattani, Chiang Mai, Chachoengsao, Chaiyaphum, and Surin. No suppression token is observed; a Sakaeo registry/ETL level break is flagged. Six candidate joins are audited. The national MMR2 panel has only two unequal-duration waves and no corresponding seventy-seven-province autism panel. Chachoengsao, Tak, and Satun matches collapse to one province, incompatible geography, or no policy/control variation. Fitting a fixed-effects or event model would create no counterfactual, so the source rows are retained without a pseudo-causal regression.

Botswana, Jordan, and Malaysia. Botswana pre-primary school censuses provide 300 region-category rows for ten education regions in 2018 and 2022; exact autism totals rise from five to ten. Secondary publications provide 210 national rows, with autism totals four, five, six, and ten in 2018, 2019, 2020, and 2022. Counts are too sparse and vaccination exposure is not available at matching region/timing. Jordan's 2023 survey supplies twelve governorate vaccination rows; its available 2015 outcome supplies eighty-four governorate broad functional-difficulty rows, not autism. Malaysia's official catalog exposes exact routes for 2018 state MMR coverage and 2018-2019 state special-education disability-category workbooks, but the archive resource downloads repeatedly time out. Even if recovered, the exposure is one wave. The audit records the routes and block without treating unavailable values as zero.

France CNAM pathology extracts. Scripts 289-292 examine the complete seventy-nine-line official pathology taxonomy and use filtered official exports rather than downloading the 5.796-million-row full file. Six claim-relevant families yield 408,240 normalized cells for 2015-2023 across 101 departments, eighteen regions, national totals, twenty-one age categories, and three sex strata. Sixty-five thousand forty-four cells suppressed below eleven are retained missing; population denominators are complete. The closest childhood psychiatric algorithm includes F84 but pools it with F80-F83, F88-F95, and F98. It is not autism-specific. The 21,816 cells overlapping the other France panel match exactly and are same-source redundancy, not independent replication.

France model and power audit. Of 211 successful descriptive models, 169 are non-placebo and forty-two are future-younger-cohort placebos. Thirty-seven non-placebo rows are nominally significant; two survive BH adjustment, both protective-sign DTP3 associations for the broad age-zero-to-four psychiatric algorithm. The female TWFE estimate is RR 0.676 (0.589-0.776) and all-sex is 0.760 (0.666-0.866), but department-trend versions are 0.932 (p=.499) and 0.869 (p=.229). Eight of forty-two future placebos reject nominally and none after BH. Fifty-one of 169 non-placebo models have at least eighty-percent power for RR 1.20 per ten coverage points, ten for RR 1.10, and the median eighty-percent MDE is RR 1.274. Precision therefore documents what the descriptive regression could detect conditional on its model; it does not cure broad outcome definition, ecological exposure, residual trends, or unavailable future placebos for many age-zero-to-four specifications.

Coefficient, power, and causal-gate ledgers. File 1481 contains 571 standardized rows: 180 British Columbia school-reconstructed screens, 180 native-district geography robustness screens, 169 France non-placebo models, and forty-two France future-placebo models. Every row carries its native scale, confidence interval, MDE where available, ecological constraint, assumptions needed for an individual mapping, binding identification constraint, source file, and zero primary causal weight. File 1482 summarizes design-specific power; file 1483 records that none of the fifteen jurisdiction or source systems passes the primary causal gate; file 1484 validates the consolidation. These rows are displayed separately rather than pooled with the primary MMR designs.

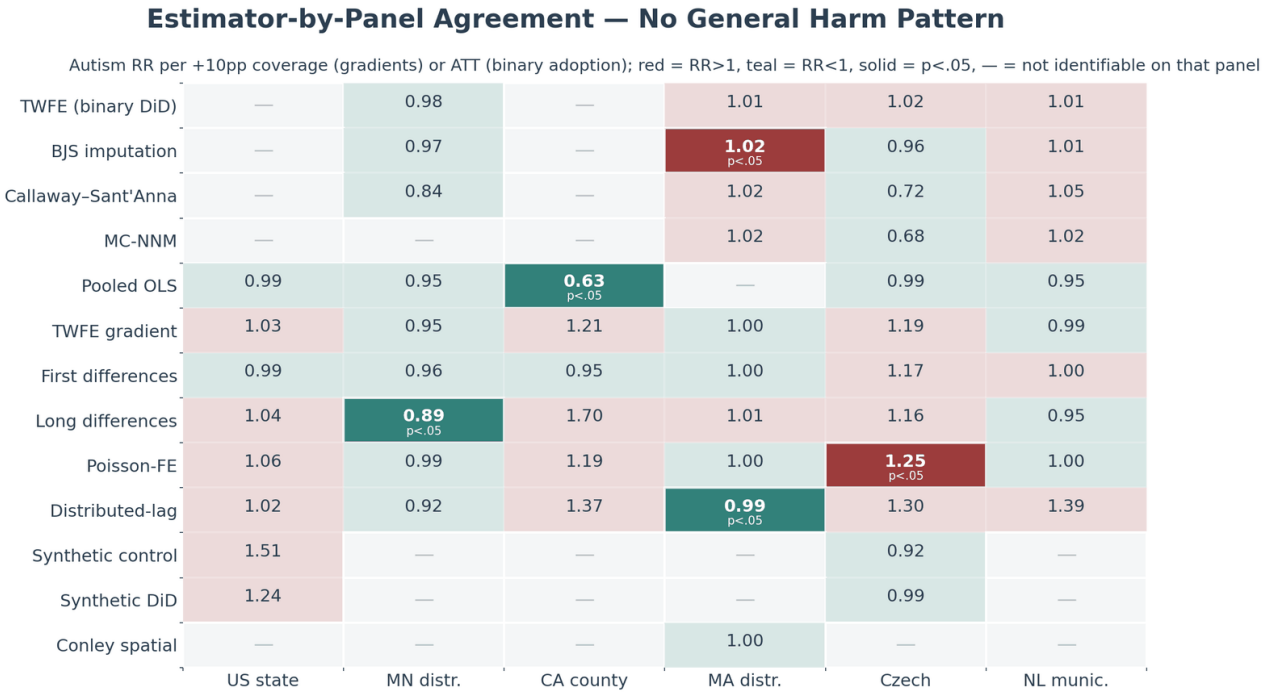
Stopping rule. The source audit closes only the unattended public-data phase after each candidate is assigned an official route, acquisition status, smallest usable unit, outcome compatibility, and next action or binding block. Remaining high-value gains require custodial person-level linkage, agency custom tabulation, historical school/district-to-health-area crosswalks, or publication of currently authenticated or withheld tables. The paper does not claim that no obscure local file exists anywhere in the world.

S22. Trend-invariant kinks, estimator-by-panel agreement, and synthesis guardrails

Trend-invariant introduction estimand. A pooled deviation-from-pretrend event study is not identified independently of its counterfactual trend: the three-cohort post-introduction coefficient is negative under a linear pretrend and positive under quadratic or level-normalized specifications. The synthesis therefore excludes a deviation coefficient and instead estimates the local cohort kink, the second difference in log outcome at the first-exposed cohort, which removes levels and linear cohort trends. The overlap-safe pool prespecifies age and retains four mutually exclusive spines - California CASEMIS, Minnesota, New Jersey, and Japan - giving RR 0.983 (95 percent CI 0.928-1.043; p=.429). Japan supplies 53.9 percent of inverse-variance information and the effective inverse-variance design count is 2.34. Removing Japan gives a U.S.-substate-only RR of 0.962 (0.857-1.081), still below 1.10 at the specified ages. Across thirty alternative fixed-age combinations its upper limit ranges from 1.072 to 1.115 and fails to exclude 1.10 in nine; with Japan all thirty exclude it. Before pooling, Japan's standard error is multiplied by 3.14, calibrated from eleven well-supported non-event cohorts. Federal IDEA is not combined with its constituent California, Minnesota, and New Jersey populations, and age is not selected by minimum standard error. The broader federal-IDEA-plus-Japan diagnostic is 0.986 (0.765-1.271). The 1999-2000 PCV7 and thimerosal boundary produces equal-and-opposite adjacent kinks in the same systems, identifying a one-cohort level blip rather than a protective PCV effect.

State-specific calibration. The EDFacts component is rerun separately in all fifty states and the District of Columbia. The pooled PCV7 composition kink is -0.00945 (clustered SE 0.00940), RR 0.991 (0.973-1.009). State z statistics remain approximately between -1.7 and +1.7. Across non-event cohort kinks the mean absolute z statistic is 0.80, the half-normal benchmark under correctly calibrated null inference. These fifty-one rows are a heterogeneity and standard-error audit of the same federal source, not fifty-one independent studies and not additional weight in a meta-analysis.

Estimator-by-panel audit. Thirteen estimator families are attempted across the U.S. state, Minnesota district, California county, Massachusetts district, Czech okres, and Dutch municipality panels. Fifty-five unique family-panel cells are identified. The long output contains eighty-five rows because some cells have multiple fitted variants; the display uses the design-preferred first implementation for each cell. Two cells are nominally positive: Czech Poisson fixed effects, RR 1.245 (p=.040), and Massachusetts BJS imputation, RR 1.021 (p=.026). Three are nominally inverse. The large California pooled-OLS inverse cell, RR 0.63, is a no-fixed-effects cross-section and moves to RR 1.21 under two-way fixed effects. Neither positive cell reproduces within its panel's remaining estimators. Eight exemption-repeal synthetic-control or synthetic-DiD rows are numerically estimable but causally ineligible: their first stages range from -1.263 to +4.0 percentage points, the outcomes are broad ages-six-to-eleven administrative stocks, and no exchangeability-dependent treated-unit placebo rejects. They receive zero causal weight. These counts are descriptive dependence-aware diagnostics, not independent Bernoulli trials. For scale, an independent staggered difference-in-differences analysis of CDC kindergarten data associated total repeal of nonmedical exemptions (California 2015, New York 2019, Maine 2019, Connecticut 2021) with a 4.0-percentage-point rise in kindergarten MMR coverage (95 percent CI 3.1-4.9) within three years (Bald, Gold, and Yang 2026); even correctly signed exemption shocks are coverage contrasts of about four points or less.



55 identified (estimator, panel) cells across the US-state reference and five newly-covered panels in North America and Europe, plus a synthetic-control / SDID audit of US exemption repeals (0 of 8 reject under Abadie placebo permutation). Two cells are nominally significant in the harm direction (Czech Poisson-FE and Massachusetts BJS); both are isolated within their own panel's battery and un-replicated across panels; the credible heterogeneity-robust suite (BJS / Callaway–Sant'Anna / MC-NNM) is null everywhere else, and the one large protective cell (California pooled OLS 0.63) is a no-fixed-effects cross-section that vanishes under its own two-way FE (1.21).

Figure S33. Estimator-by-panel agreement across six public panels

Conditional power and Bayesian evidence. The power figure reports model-conditional 80-percent MDEs, power at RR 1.10, and BF01 under a positive half-normal harm prior with scale log(1.2). The first-difference and thirty-seven-country fitted models yield BF01 31.4 and 44.5. Those Bayes factors quantify likelihood evidence conditional on each specification and prior; they cannot repair an ecological estimand, a failed pretrend, broad outcome definition, or dependence among source pipelines. The trend-sensitive deviation event study remains in the display only as an explicit nonidentified diagnostic and carries no introduction claim. The introduction claim is evaluated by the kink estimator above.

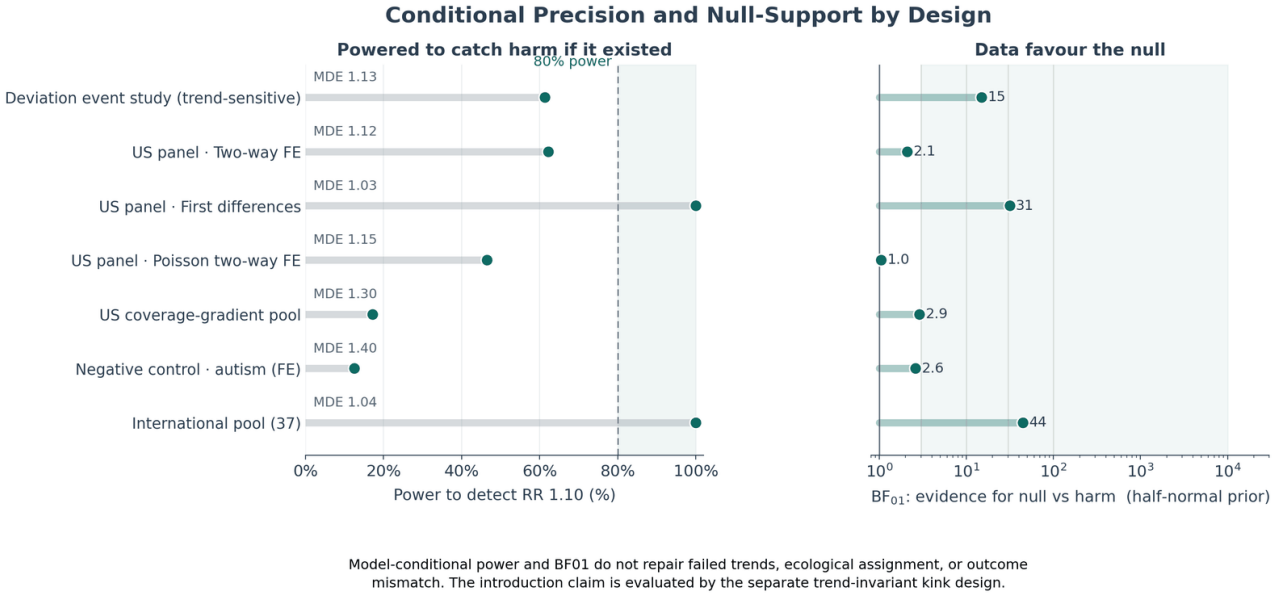


Figure S34. Design-conditional power and Bayes factors under a positive half-normal alternative

Full state-panel multiverse. The expanded curve contains 156 temporally ordered specifications spanning OLS, Poisson, first differences, long differences, several samples, lags zero through two, trend choices, weights, and HC1 versus state-clustered inference. All fitted RRs lie between 0.978 and 1.102. Thirteen nominal harm-direction results use HC1, which is anti-conservative under serial dependence; zero of seventy-six state-clustered specifications rejects. The thirteen HC1 results also fail familywise Holm and Benjamini-Hochberg adjustment. The multiverse is evidence about inference sensitivity within one panel, not 156 independent studies.

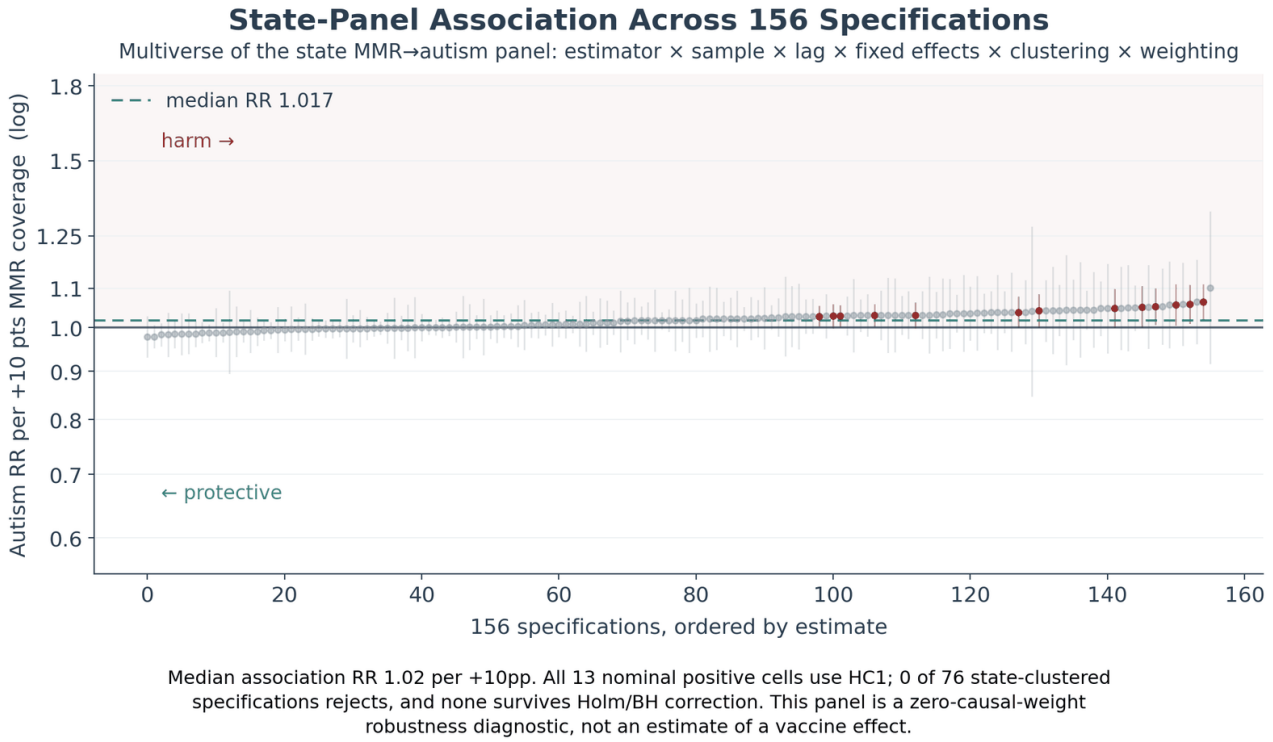


Figure S35. Full 156-specification state-panel multiverse and inference audit

The nine-design value RR 1.001 (0.993-1.008) in file 150 combines unlike contrasts, including failed-pretrend national steps from Northern Ireland and Sweden. It is not a causal effect or an individual-risk bound. Neither this aggregation nor the archived PCV7 fixed-share calibrations supply current individual estimates. S53 reports full exposure-model analyses without substituting nominal shares for a fitted rollout contrast.

S23. Dependence, first-stage uncertainty, and multiplicity

Generated exposures. The PCV7 and RotaShield state intensities are survey estimates rather than fixed regressors. File 1890 redraws each state-year coverage cell from a Jeffreys-smoothed beta distribution using its Kish effective sample size, rebuilds the state intensity, and refits the full Poisson model 249 times. The resulting first-stage standard errors are 0.00340 for PCV7 and 0.00871 for RotaShield. Adding these variances conservatively to the archived state-cluster variances yields PCV7 RR 1.013 (0.981-1.046) and RotaShield RR 0.996 (0.945-1.051). Approximate classical reliabilities are 0.971 and 0.907. Exposure sampling error is therefore real but not the explanation for either null.

Serial and spatial dependence. The state MMR coefficient is RR 1.034 per ten coverage points. It is null with state clustering (0.953-1.122), two-way state-year clustering (0.963-1.111), and every archived Conley spatial-temporal covariance (the narrowest interval is 0.991-1.079). Driscoll-Kraay bandwidths one through three instead give 1.016-1.053, 1.017-1.052, and 1.018-1.051. This disagreement is reported rather than resolved by choosing the smallest standard error. With only seventeen irregular calendar waves and strong common time adjustment, the cross-sectional large-N asymptotic behind that covariance is less credible than the state-cluster, two-way, and spatial checks; the association remains a secondary ecological gradient in every case.

Trend-invariant and small-cluster tests. Northern Ireland second-differences the cohort effects, which removes the unidentified linear APC drift. Because the curvature regressor is generated from the estimated age-period-cohort baseline and there are few report-year clusters, inference jointly refits the baseline and curvature in a report-year cluster bootstrap. The resulting RRs are 1.065 (0.952-1.163) for rotavirus, 1.081 (0.948-1.222) for MenB, and 0.986 (0.822-1.413) for 6-in-1. Tiny HC/Newey-West MenB p-values are classified as generated-regressor/few-cluster artifacts, not evidence. The observed kinks also have empirical non-event rank p-values 0.33, 0.52, and 0.33. Sweden's F84 PCV model gives RR 1.486 (t20 interval 0.840-2.628) and a 999-draw 21-region wild-cluster p-value of 0.162. Second differences of its normalized dynamic profile are non-significant at every event center. These tests extract valid diagnostics from failed normalized pretrends; they do not convert the national or care-utilization panels into valid level-effect designs.

Pooling and multiplicity. The canonical trend-invariant synthesis is the overlap-safe, specified-age four-spine pool of California CASEMIS, Minnesota, New Jersey, and Japan: RR 0.983 (0.928-1.043; p=.429). Japan supplies the largest weight (53.9 percent), the effective design count is 2.34, and the leave-Japan-out U.S.-substate sensitivity is 0.962 (0.857-1.081). Federal IDEA is excluded from the canonical pool because it contains the U.S. substate populations; IDEA-plus-Japan and the overlapping five-spine covariance-engine pool are separate diagnostics. BH and Holm corrections across 835 coefficient screens give eight BH and seven Holm rejections, all with zero positive primary-causal-pool weight. Statistical multiplicity cannot promote a design that fails its cohort, geography, outcome, or counterfactual gate.

Additional official-source results. Poland's 2024 Podlaskie sanitary report yields ninety powiat-by-antigen coverage rows for children in their third year of life: fourteen powiats plus the voivodeship, six antigens, and an MMR range of 80.7-95.0 percent. It is retained as a one-wave exposure component because no concordant repeated powiat named-autism outcome was verified. South Africa's public District Health Barometer endpoint yields 90,807 normalized observations, fifty-six health-system indicators, and district plus smaller-municipality geographies for 2018-2024. The eligible district screen fits 324 log-outcome models for under-one immunization and measles-dose-two coverage at future, contemporaneous, and lag-one timings with district and year fixed effects. District-clustered inference gives thirty-six nominal and six within-family BH rejections, all service co-performance measures; district-and-year clustering gives twenty-three nominal and zero BH rejections. These rows have zero causal weight because there is no cohort alignment, schedule shock, individual exposure, or named-autism outcome. Files 1900-1907 preserve the raw hashes, full panel, every coefficient, MDE, source gate, and validation result. Chile, Malaysia, and Costa Rica do not provide concordant public causal pairs; unobserved fields are not coded as zero.

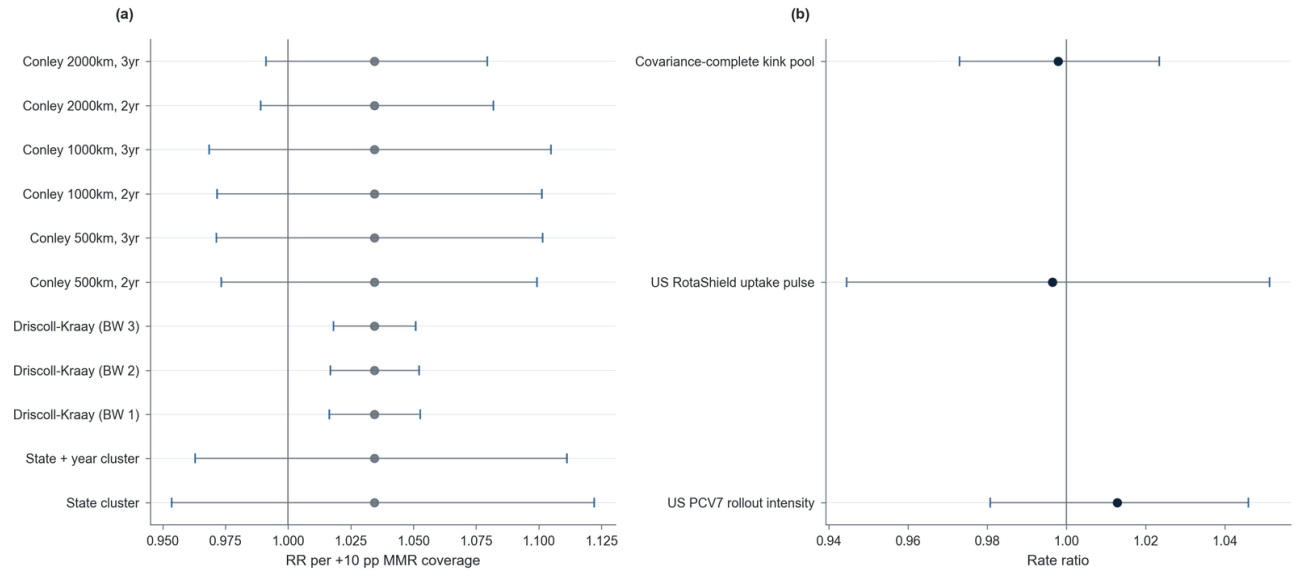


Figure S36. Dependence-robust and uncertainty-complete focal estimates

S24. Exact small-k, multivariate-outcome, specification-curve, and suppression-envelope inference

Exact sign diagnostic for the historical local mixture. Enumerating all 512 sign assignments of the nine jurisdiction scores gives two-sided p = .352; the fixed-effect ratio is 0.992. Wisconsin contributes 65.1 percent of the inverse-variance weight, but its exposure is complete-schedule compliance, not MMR. Five other rows have special-education denominators and three have enrollment denominators. The unweighted median is 0.992 (distribution-free 96.1 percent interval 0.990-1.022), and leave-one-out fixed summaries range 0.991-0.996. These are descriptive statistics of the heterogeneous diagnostic set, not nine replications of a common MMR effect. S67 supplies the exposure- and denominator-matched analyses; file 2052 preserves this calculation.

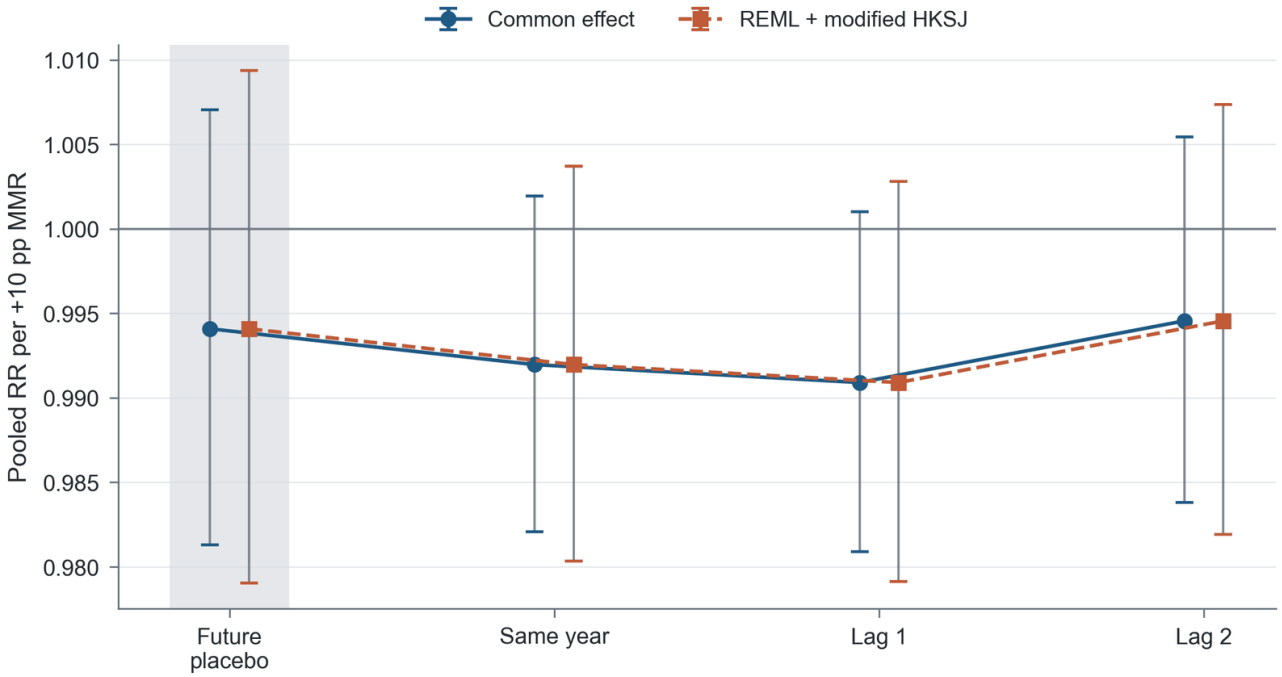


Figure S72. Local MMR contemporaneous, lagged and future-exposure estimates. The figure retains one publication pipeline per jurisdiction. These are ecological associations on the stated outcome denominators, not individual vaccination risks; alternative timing specifications reuse observations.

Dependence-correct specification curve. Fifty-six linear state-cluster-appropriate specifications are jointly calibrated with common state-level wild weights. The median is RR 1.018, two-sided wild p=.320. The maximum positive t statistic is 1.313 with familywise p=.684; the simultaneous critical value is 3.56 and zero of fifty-six simultaneous intervals excludes one. Archived coefficients and clustered standard errors reproduce to numerical tolerance (file 2053).

Multivariate outcomes. Joint wild inference across autism and seven IDEA categories rejects the nondirectional all-zero coefficient vector (Q=19.45, p=.002), reflecting heterogeneous category movements rather than generalized injury. The maximum positive coefficient has familywise p=.606, the equal-weight outcome mean is RR 0.988 with one-sided harm p=.677, autism versus the sensory/motor mean is RR 1.053 (p=.135), and autism versus all comparators is RR 1.029 (p=.217). No individual outcome rejects after max-absolute-t correction. These are jointly modeled claim-relevant outcomes, not a relabeling of every non-autism condition as a negative control (file 2051).

Suppression scenarios. Crossing 243 fitted lower/midpoint/upper endpoint combinations for California, Illinois, Pennsylvania, and Oregon gives lag-zero pooled point RRs 0.9907-0.9933 and a union of conventional intervals 0.9798-1.0035. No scenario has a lag-zero point estimate above one or an interval wholly above one. This is a conservative stress envelope across published masking choices, not sharp partial identification over all cellwise latent counts (file 2054). File 2055 records why augmented synthetic control, confirmatory nine-unit meta-regression, further spatial HAC, and a prior-driven bias model would not repair the design limits; file 2056 provides the corresponding validation results.

S25. Residual methods and public-data claims

National varicella diagnostic. The national single-age IDEA/Census cohort analysis covers U.S. varicella licensure and recommendation. Age and reporting-year fixed-effects models on birth cohorts 1982-2000 give a 1995 post-cohort RR of 0.870 (0.791-0.958); the step-plus-kink level term is 0.882 (0.809-0.962) and its hinge RR is 0.925 per cohort. This apparently protective result is not a regression discontinuity and receives zero causal weight. Licensure occurred in 1995, routine recommendation in 1996, catch-up and gradual uptake blur assignment, and the actual-cutoff absolute t statistic ranks only p=.40 among admissible 1987-1996 placebo cutoffs. Files 2071-2072 classify the pattern as broad cohort curvature, not evidence of protection.

Positive-weight dependence in the historical local mixture. Holding the nine positive inverse-variance weights fixed gives ratio 0.992; its t8 interval widens from 0.980-1.004 at assumed pairwise sampling correlation zero to 0.968-1.016 at correlation one. The effective inverse-variance design count is 2.11 and Wisconsin carries 65.1 percent of the weight. Wisconsin's county-cluster interval is 0.979-1.002 around ratio 0.990, versus 0.978-1.002 under district clustering. These calculations (files 2073, 2073b and 2075) describe dependence sensitivity of a mixed-denominator, mixed-exposure diagnostic. They do not correct its estimand incompatibility, validate an MMR contrast or establish an individual-risk bound. S67 reports separate native families with explicit shared-bias sensitivities.

Conditional sharp suppression sets. The 243-scenario jurisdiction-level envelope moves whole endpoints together and is not sharp. Cellwise linear optimization of the dominant outcome rectangles, conditional on midpoint exposure and the TWFE design matrix, gives RR sets 0.801-1.212 for Wisconsin and 0.520-1.871 for California. Holding the other seven estimates and positive pooling weights fixed gives a dominant-two conditional range 0.757-1.289. These sets are sharp only for the specified outcome rectangles; they do not cover exposure masking, sampling uncertainty, other jurisdictions' latent cells, or shared bias. They therefore prevent the uniform-endpoint envelope from being overinterpreted without turning an incomplete bound into a global identified set (file 2074).

Feasibility adjudication and review status. Public NHIS and NSCH files select a focal/sample child and do not jointly contain sibling-specific verified vaccine histories plus ASD outcomes, so sibling fixed effects cannot be estimated. No held exposure-outcome file combines subannual eligibility timing, a measured vaccination first stage, and the exclusion restriction needed for fuzzy RD. E-values, Oster proportional-selection analysis, HonestDiD, synthetic controls, synthetic DiD, and matrix completion were already implemented where their inputs exist. The source review found no additional immediately downloadable compatible causal-ready exposure-autism pair; DHS/MICS requires registration and generally measures functioning rather than diagnosed autism, while historical California DDS county-by-cohort cells are request-only and still lack a verified matching exposure panel. The adjudication supports the dependence accounting and narrower family-specific conclusions with reservations; it does not certify that all possible global data have been obtained. File 2076 records every suggestion and disposition; file 2077 passes all eleven adjudication checks.

S26. Analytical decision protocol and hierarchy of evidence

Purpose. No observational or quasi-experimental choice is assumption-free, and the supplement does not describe any estimator as unimpeachable. The analytical protocol instead separates four questions that are often collapsed: whether an exposure change is real, whether the outcome measures the proposed condition, whether a counterfactual can be learned, and whether the resulting interval is precise enough to constrain the claim. A small p-value cannot repair a failed exposure, outcome, or counterfactual gate. A null estimate cannot establish absence when its confidence interval or design-based MDE is wide. These rules are applied before direction is considered.

Six-gate sequence. Each design is evaluated on: (1) exposure provenance and size; (2) cohort alignment, including administration age and transition cohorts; (3) outcome definition and reporting-system stability; (4) counterfactual support, including untreated units, differential uptake, or a valid sign-reversing withdrawal; (5) timing diagnostics, pretrends, future-exposure placebos, and non-event cutoffs; and (6) inference and power at the independent assignment level. Designs passing all six may enter a narrowly defined synthesis family. Designs failing one or more remain in the register and figures but receive zero primary-pool weight. This is a design gate, not a significance filter: harmful, protective, and null estimates are downgraded by the same rules.

Evidence roles. The project distinguishes primary quasi-experimental estimates, within-family robustness estimates, descriptive ecological associations, individual observational calibration, exposure-only and outcome-only components, and access or feasibility records. The historical register contains 737 records, the associated source catalog contains 400 records, and its statistical-result ledger contains 25,559 rows; these are distinct record layers, not study counts. S70 reports the full integrated result register. Observational calibration receives zero causal weight. A source can be valuable for confirming a first stage, reproducing a public total, checking a reporting pipeline, or documenting a missing join even when it contributes no causal coefficient. The all-results accounting assigns every raw specification row zero omnibus pooling weight.

Choice timing. Whenever several estimators are feasible, the default is defined by the assignment process and data geometry rather than by fit or sign. Continuous uptake gradients use unit and cohort or year fixed effects with clustering at the assignment unit. Staggered binary adoption uses cohort-specific group-time, imputation, or stacked estimators when untreated support exists. A single treated unit with credible donors uses synthetic-control methods. A national change with no untreated geography is not relabeled staggered DiD; it is treated as a weaker cohort-curvature, interrupted-series, or sign-reversal exercise. Preferred ages, lags, and exposure definitions are fixed by schedule timing and ascertainment, not by the minimum standard error.

Escalation rule. A result becomes more credible when it survives a specified ladder: raw paths; fixed effects; alternative links; flexible trends; one-step-up trend complexity; preperiod-only path selection; future and fake-event placebos; alternative estimators; influence and leave-one-unit checks; masking and measurement-error sensitivity; serial, spatial, and common-shock inference; and familywise multiplicity correction. Failure at an early identification step is retained and labeled. Later robustness cannot turn a nonidentified national step into a causal design.

S27. Cohort construction, exposure assignment, and outcome definitions

Core index. The preferred analytical cell is geography g by birth cohort b, observed at a fixed ascertainment age a. For exact-age administrative outcomes published in reporting year t, cohort is constructed as b = t - a. Exposure is assigned to the same cohort using the scheduled administration age and the coverage statistic's age cutoff. Calendar-year conventions are audited with plus-or-minus-one-year shifts. When only broad age bands are published, the design is labeled a cohort-stock or composition analysis rather than an exact-cohort prevalence design.

Transition cohorts. A policy year is not automatically the first fully exposed cohort. For a programme beginning during calendar year y, births in y may be partially eligible, may receive catch-up, or may be measured before completing the series. The analysis records the legal start, operational start, scheduled dose age, first fully exposed annual cohort, catch-up rule, and last unexposed cohort. Transition cohorts are excluded from sharp nested-state comparisons and retained only in continuous-share or partial-exposure specifications. This is especially important for English rotavirus, MenB, and 6-in-1 revisions and for the Japanese decline that preceded formal MMR suspension.

Coverage measures. Provider-verified survey coverage is preferred to parental recall. Count and denominator pairs are retained whenever published; percentages are never treated as counts. Dose completion and antigen components remain separate. A combined schedule-status measure is not used to identify a pairwise vaccine interaction. When

exposure is estimated, its sampling variance is propagated by bootstrap or combined-variance sensitivity when the design permits; otherwise attenuation is discussed and reliability sensitivity is reported.

Outcome hierarchy. Clinical cumulative incidence, clinician-reviewed prevalence, health-care utilization, education-service classification, disability composition, parent-reported diagnosis, and broad functioning measures are not interchangeable. Analyses preserve the native outcome and denominator. IDEA autism is an education-service category, ADDM is a clinician-reviewed surveillance construct, specialist-care F84 counts utilization, and MICS child functioning is not diagnosed autism. Outcome-system differences are reasons for separate families, not labels to be erased by rescaling.

Reporting changes. Category introductions, coding revisions, diagnostic-instrument transitions, geographic reorganizations, and source breaks are encoded as explicit events or sample restrictions. Raw special-education totals before and after an autism category is introduced are not interpreted as incidence. The design checks whether a claimed exposure effect coincides instead with a reporting discontinuity. Alternative outcomes such as autism plus speech/language impairment and autism share of all IDEA categories are used to test substitution, not to manufacture a preferred denominator.

Geographic joins. School, district, county, health authority, region, and state crosswalks are versioned by year when possible. Exact identifiers dominate fuzzy name matching; spatial joins retain match distance and polygon status. A national exposure cannot be assigned to local outcomes as if it supplied local treatment variation. If exposure and outcome geographies do not nest, the source remains a descriptive component unless a defensible population-weighted crosswalk is constructed and sensitivity to that mapping is shown.

S28. Estimands and estimating equations

Continuous uptake model. For count outcomes, the main ecological gradient is a log-link model with population offset:

$$\log \mathbb{E}[Y_{gb}] = \alpha_g + \lambda_b + \beta X_{gb} + \mathbf{Z}'_{gb}\gamma + \log N_{gb}$$

Here X is coverage or rollout intensity in a stated unit, usually ten percentage points or one rollout standard deviation; N is the relevant population or enrolment denominator. The reported RR is exp(beta). Linear probability or rate models are fitted as functional-form sensitivities because parallel trends in levels and logs are distinct assumptions.

Event-study model. Dynamic models replace X with event-time indicators interacted with exposure intensity or treatment cohort. A reference preperiod is omitted, and leads are displayed jointly rather than individually screened. The estimand is a differential outcome path, not a simple national time break. When adoption is staggered, cohort-specific group-time, imputation, Sun-Abraham, or stacked comparisons are used where untreated support exists. When all units change together, unit fixed effects cannot identify the national level step separately from cohort time; only differential intensity, slope curvature, or other restrictions create an estimand.

Difference-in-differences and DDD. The primary DiD contrast compares changes across units with different exposure changes, conditional on unit and cohort or year effects. Diagnostic-substitution DDD adds an outcome-category dimension and asks whether autism moves relative to claim-relevant comparison categories in the same reporting system. A DDD can reduce common ascertainment changes but introduces the assumption that the comparison category is not itself affected by the exposure. Because vaccine-critical claims extend to many conditions, those categories are described as claim-relevant outcomes rather than universally valid negative controls.

Withdrawals and pulses. A withdrawal has a sign prediction: if an exposure materially raises later autism, a large removal should lower later identification among affected cohorts. Yokohama is modeled with its continuous cohort coverage path because exposure declined before the legal suspension. RotaShield is modeled as a measured low-uptake state pulse rather than a national binary indicator. Introduction and withdrawal components are not pooled unless their effect units and timing correspond.

Schedule states and interactions. Pairwise coverage models include centered main effects for vaccines A and B and the product of their ten-point changes. The interaction is a departure from log additivity, not automatically a biological synergy parameter. Identification requires independent support for combinations of A and B; when products rise together, variance inflation and support diagnostics determine whether the interaction is estimable. Nested-state models

instead compare schedule A with schedule A+B to estimate the incremental change associated with adding B, exclude transition cohorts, and use future-state terms as timing placebos. That incremental contrast is not an A-by-B interaction unless the relevant off-diagonal schedule states are also supported. A combined schedule-compliance measure cannot identify antigen-specific synergy.

A null interaction is not a null cumulative effect. For risks in the four states neither, A only, B only and A+B, additive interaction subtracts the two single-product risk increments from the combined increment. Multiplicative interaction instead divides the A+B versus neither risk ratio by the product of the two single-product risk ratios. The following invented probabilities demonstrate the distinction; they are not observed vaccine risks.

Synthetic example	Neither	A only	B only	A+B	A+B versus neither RR
No additive interaction	1.00%	1.10%	1.10%	1.20%	1.20
No multiplicative interaction	1.00%	1.10%	1.10%	1.21%	1.21

The first example has zero excess risk on the additive interaction scale despite a 20% combined increase; the second has interaction ratio one despite a 21% combined increase. Thus, an interval around one for a log-link interaction term cannot itself exclude those main-effect-only cumulative models. Joint main effects, baseline risks, exposure support and identification assumptions are needed to bound a combined-versus-neither contrast. Changing the interaction scale can also change whether excess interaction is present. Script `check_interaction_estimands.py` verifies these arithmetic counterexamples; they contribute no empirical observations or evidence weight.

The held PCV interaction has 252 state-cohort rows across fifty-one jurisdictions and five proxy birth cohorts, 1999-2003. Its baseline variable is the arithmetic mean of MMR and varicella receipt percentages, not total vaccine doses or the fraction receiving both products. PCV coverage ranges from 10.15% to 99.54%, and the baseline mean from 73.27% to 98.09%. Neither the zero-coverage pair nor the universal-coverage pair lies inside the pooled exposure convex hull. Among observed points used as lower-left anchors, forty support all four corners of a simultaneous ten-point coverage rectangle inside the pooled hull. Within-cohort counts are five, six, zero, zero and zero. These counts describe a particular positive-shift interpolation diagnostic, not a test of causal positivity or a necessary condition for algebraic identification of a global interaction coefficient. Hull membership also does not establish nearby observation density, within-state support or exchangeability. Monotone-chain and independent Qhull calculations agree for the pooled and five cohort hulls.

In the centered log-link model, the joint increase in both exposures depends on both main-effect coefficients, the interaction coefficient and the starting exposure levels. By contrast, the four-cell mixed difference of log means isolates the interaction. Converting its reported interval directly into a combined-versus-unvaccinated bound would omit those main effects and extrapolate outside the observed coverage support. Individual joint-receipt patterns used to propagate NIS sampling uncertainty do not link the IDEA outcome records to those same children. The support and algebra checks are recorded by `audit_interaction_exposure_support.py` and `verify_interaction_exposure_support.py`; they do not add a causal estimate.

Main-effect-inclusive contrasts use the same Poisson model and its full coefficient covariance. For each five- or ten-percentage-point shift, the comparison anchors are observed state-cohort points whose four shifted coverage corners lie inside their own cohort's exposure hull. There are seventy-six five-point anchors and eleven ten-point anchors. We report geometric averages of conditional rate ratios across those anchors, holding population offsets and state/cohort effects fixed. These are different target subsets, not a dose-response series. Independent GLM and score-based calculations agree on the coefficients and state-cluster covariance.

Coverage change	Contrast	Conditional ratio	Pointwise 95% interval	Eight-contrast 95% interval
+5 points	PCV only	1.000	0.981-1.019	0.973-1.027
+5 points	Mean MMR/varicella only	1.000	0.962-1.039	0.947-1.055
+5 points each	Both coverage measures	0.999	0.963-1.037	0.949-1.053
+5 points each	Excess multiplicative interaction	1.000	0.997-1.003	0.996-1.004
+10 points	PCV only	0.999	0.962-1.037	0.947-1.053
+10 points	Mean MMR/varicella only	0.998	0.920-1.083	0.889-1.121
+10 points each	Both coverage measures	0.998	0.925-1.076	0.896-1.111
+10 points each	Excess multiplicative interaction	1.001	0.990-1.011	0.985-1.016

The intervals condition on published exposure regressors, using state-cluster CR1 covariance and a t reference with fifty degrees of freedom. The final column applies a Bonferroni adjustment to these eight contrasts only, not to the preceding research search, and requires valid component intervals. No exposure-redraw, survey-design or support-selection uncertainty is included here; these intervals must not replace the exposure-adjusted interaction results elsewhere. Shared shocks, confounding, limited follow-up and outcome linkage remain unresolved. Joint coverage contrasts are appreciably less precise than the interaction term alone. Neither their near-one point estimates nor the geometric support check identifies a total-vaccine-load or vaccinated-versus-unvaccinated causal effect. Scripts `estimate_supported_joint_coverage.py` and `verify_supported_joint_coverage.py` retain all eight contrasts and verify the changed-feature arithmetic and covariance propagation.

A separate sensitivity refits the model in 499 joint exposure redraws, using the eight NIS receipt-pattern probabilities with Dirichlet concentration equal to Kish effective sample size times each probability plus one-half. This preserves joint receipt covariance approximately; it is not a complete survey-design variance procedure. Original raw-percentage targets and anchor identities remain fixed, while centering follows each refitted design. No draw or target is dropped because a support check fails. The reported centers exponentiate the mean log contrast. Sensitivity intervals combine mean within-draw state-cluster variance with between-draw log-contrast variance and use the same t multipliers as above. They have no demonstrated frequentist coverage and are not a calibrated posterior.

Coverage change	Contrast	Redraw center	Pointwise sensitivity interval	Eight-contrast sensitivity interval
+5 points	PCV only	1.000	0.983-1.017	0.976-1.025
+5 points	Mean MMR/varicella only	1.001	0.969-1.033	0.956-1.047
+5 points each	Both coverage measures	1.001	0.970-1.033	0.957-1.047
+5 points each	Excess multiplicative interaction	1.000	0.997-1.003	0.996-1.005
+10 points	PCV only	1.000	0.966-1.034	0.953-1.049
+10 points	Mean MMR/varicella only	0.999	0.931-1.071	0.904-1.103
+10 points each	Both coverage measures	1.000	0.936-1.068	0.911-1.098
+10 points each	Excess multiplicative interaction	1.001	0.989-1.013	0.984-1.018

The number of fixed anchors retaining all four corners inside their own cohort's redrawn hull ranges from sixty-two to seventy-six of seventy-six for five-point changes (median seventy-one), and three to eleven of eleven for ten-point changes (median ten). Some refitted predictions therefore extrapolate even though their targets passed the original geometric screen. The narrower intervals for some redraw contrasts reflect changed regressors and conditional model variance, not additional observations or established precision gains; they do not supersede the fixed-exposure intervals. Support uncertainty, systematic exposure error and confounding are not resolved by this exercise. All 3,992 contrast-draw records and 998 support records are retained and their moments independently checked by `redraw_supported_joint_coverage.py` and `verify_joint_coverage_redraws.py`.

Saturation estimand. The saturation test asks whether autism identification continues rising after measured schedule expansion and uptake drift are too small to generate the trend under a specified large causal RR. It is a bounding comparison, not a regression proving no unmeasured schedule dimension changed. The calculation states the assumed

dose-response, coverage drift, and latency. Continuing outcome growth rejects the claim that the measured additions or uptake trends mechanically explain the secular rise; it does not prove every environmental cause absent.

Individual microdata. Public NHIS, NSCH, and SLAITS analyses use survey weights and design variables where available, retain vaccination-documentation indicators, and report sparse-cell counts. Their estimand is an adjusted observational association between recorded vaccination and parent-reported diagnosis. They are not used as instruments for the ecological estimates and do not enter the quasi-experimental causal pool. Sibling fixed effects are not feasible because the public files do not jointly observe sibling-specific verified histories and ASD outcomes.

S29. Estimator applicability and modern DiD alternatives

Why no universal estimator. Estimators answer different questions and require different data. TWFE is appropriate for continuous treatment gradients under additive unit and time effects but can be misleading for staggered binary adoption with heterogeneous effects. Callaway-Sant'Anna group-time ATT, Sun-Abraham interaction-weighted event studies, Borusyak-Jaravel-Spiess imputation, and stacked DiD are fitted when treatment cohorts and untreated comparisons exist. Agreement across them is reported as a robustness map; disagreement is a diagnostic, not an invitation to choose the preferred sign.

Staggered-adoption support. A policy panel must contain untreated or not-yet-treated observations for the relevant event times. Cohorts with no comparison support are omitted rather than extrapolated. Anticipation windows are defined by eligibility and administration age. Group-time weights and cohort contributions are audited. When all units are treated in the same period, these estimators are inapplicable; adding unit fixed effects does not manufacture a control group.

Synthetic control family. Synthetic control, augmented synthetic control, synthetic DiD, and matrix completion are used for state exemption events and other small-treated-unit designs only when a donor pool exists. Preperiod RMSPE, donor weights, convex-hull support, leave-one-donor sensitivity, in-space placebos, and in-time placebos are reported. A good postperiod fit after poor prefit receives zero causal weight. Augmentation may reduce interpolation bias but cannot repair absence of donors or a national exposure.

Continuous-treatment DiD. Continuous uptake contrasts require a parallel response surface rather than merely parallel binary-treatment trends. The project reports residual exposure variation after fixed effects, within-unit exposure SD, and the contrast over which the RR is displayed. A ten-point coefficient can be a substantial extrapolation when residualized coverage varies by only one or two points; the corresponding one-within-SD MDE is therefore also reported.

First differences and long differences. Differencing removes time-invariant unit effects without assuming the level fixed-effect functional form. First differences emphasize adjacent waves and can amplify measurement error; long differences trade additional temporal confounding for larger first stages. Both are treated as alternative estimators rather than independent evidence. Distributed-lag and lagged-outcome models are similarly dependent specifications on the same data spine.

Poisson and linear links. Poisson quasi-maximum-likelihood models with an offset retain count scale and are robust to variance misspecification under a correct conditional mean. WLS log-rate models make the ratio interpretation transparent. Identity-link rate differences reveal whether conclusions depend on multiplicative scale. A link change that moves the estimate but not the qualitative interval is reported as scale sensitivity; a sign reversal is an identification warning.

S30. Cohort trends, pretrends, and time-placebo diagnostics

Trend hierarchy. The baseline controls for common cohort or year shocks. Sensitivity models add unit-specific linear trends, quadratic trends where support is adequate, and restricted cubic splines with knots fixed before examining the outcome. Flexible splines are valuable when long smooth cohort changes are plausible, but with few cohorts they can absorb the event and create extrapolation instability. The analysis therefore reports design-matrix rank, residual degrees of freedom, observations per parameter, leverage, and the number of preperiods supporting each trend.

One-step-up rule. The preferred trend specification is accompanied by a model one level more flexible. If the effect disappears, reverses, or moves to an impossible lead under a modest increase in flexibility, the design is not described as identified. Preperiod-only selection compares candidate paths using leave-one-preperiod-out prediction; postperiod

outcomes never choose the counterfactual model. Designs with fewer than three informative preperiods cannot support this procedure and are labeled accordingly.

Pretrend testing. A nonsignificant lead is not proof of parallel trends. Leads are tested jointly, plotted with simultaneous bands, and interpreted with their power. Roth-style pretest concerns are addressed by retaining failed designs and by applying Rambachan-Roth HonestDiD sensitivity to bounded post-treatment departures where the long-lead design and software geometry permit. Designs with visibly different prepaths are not rescued by simply controlling for a fitted pretrend unless the extrapolation is stable across functional forms.

Curvature estimands. In national age-period-cohort panels, level effects can be unidentified because cohort equals period minus age. First or second differences can remove linear APC drift and identify curvature, but the estimand becomes a slope change rather than a level shift. Covariance among cohort effects is retained in the kink standard error. Mirror-pair kinks at adjacent cohorts are classified as one-year level blips. National varicella, Northern Ireland, Sweden, and other APC-restricted results are kept as curvature diagnostics unless the timing and placebo structure support more.

Placebos. The project uses future-exposure terms, younger unaffected cohorts, fake event dates, non-event cutoff ranks, alternative vaccines with no contemporaneous schedule change, permutation of exposure across units, and in-time synthetic-control placebos. State, coverage, and treated-unit labels were not randomized; permutation p-values are conditional on exchangeability of the permuted labels and are diagnostics rather than design-based randomization inference. A future placebo is interpreted as evidence of residual trend or common cause, not as a protective or harmful vaccine effect. Placebo families are calibrated jointly when derived from the same panel.

S31. Serial, spatial, small-cluster, and generated-exposure inference

Default clustering. Standard errors cluster at the level where exposure is assigned and residual shocks can remain correlated. State exposure implies state clustering; district exposure implies district clustering unless a higher shared geography is plausible. Year fixed effects absorb common additive shocks but not heterogeneous common-shock responses. Two-way unit-year clustering is reported when the number of time clusters is adequate, with explicit warning when it is not.

Few clusters. With few independent units or report years, asymptotic cluster t tests are supplemented by wild-cluster bootstrap, score sign flipping, or cluster-level permutation diagnostics. Exact enumeration can be exact for the imposed sign- or label-exchangeability null without making the observational exposure randomized. The nine-jurisdiction MMR synthesis enumerates all 2^9 score signs under sign exchangeability. Northern Ireland jointly bootstraps the generated APC curvature and outcome stage at the report-year level. Small-cluster p-values are limited by the number of possible assignments; numerical precision beyond that resolution is not reported.

Serial covariance. Newey-West and Driscoll-Kraay estimates are treated as sensitivities whose asymptotics may be weak in short, irregular panels. State-cluster and two-way cluster intervals remain primary for the flagship. Disagreement between covariance estimators is shown rather than resolved by choosing the narrowest. Autocorrelation functions and bandwidth choices are archived.

Spatial dependence. Conley covariance estimates use specified distance cutoffs and time bandwidths when unit coordinates are available. County clustering of the dominant Wisconsin equation permits arbitrary cross-district dependence within county; it does not model neighboring counties. Panels without stable coordinates or historical geography are not assigned an invented spatial kernel.

Shared-data dependence. Person-disjoint jurisdictions can share national reporting rules, vaccine-survey methods, and common diagnostic shocks. Local-MMR sensitivity holds positive weights fixed and varies pairwise sampling correlation from zero to one; a separate additive common-shock Bayesian sensitivity addresses common bias. The inverse-variance effective design count and largest weight are reported. Equicorrelation GLS rows that create negative weights are retained as diagnostics and not used as primary summaries.

Generated exposures. When rollout intensity, cohort shares, or curvature regressors are estimated, uncertainty is propagated by joint bootstrap or combined variance where feasible. Plug-in standard errors are retained only as labeled comparisons. The power calculation uses the combined exposure-outcome uncertainty rather than assuming the first stage is fixed.

Exact-model operating characteristics. A separate effect-injection audit refits the exact 353-cell, 51-state, seven-cohort PCV7 and RotaShield specifications, including the fixed effects, offsets, spline terms, residual cluster trajectories, and state-clustered covariance. These are continuous pseudo-count cluster-wild quasi-Poisson checks, not parametric count-data generators and not new identifying designs. For PCV7, the null rejection rate is 4.2 percent and empirical power is 21.8, 86.8, and 100.0 percent at RR 1.02, 1.05, and 1.10; the interpolated MDE80 is 1.0468, compared with 1.0467 after the separate generated-exposure variance layer. For RotaShield, the archived four-test Bonferroni rule uses alpha 0.0125. Its null rejection rate is 2.4 percent, above the nominal 1.25 percent by about 1.7 Monte Carlo standard errors, so its power is read conservatively. Conditional power is 4.6, 24.4, and 94.4 percent at RR 1.02, 1.05, and 1.10. Adding generated-exposure uncertainty gives MDE80 1.0959 and 83.3 percent power at RR 1.10. Files 2407-2412 archive the draws, reconciliations, and validation.

S32. Missingness, suppression, measurement error, and ecological translation

Suppression representation. Published tokens such as less than 6, less than 10, greater than 90 percent, or interval-coded percentages are stored as lower and upper bounds with the raw token preserved. Zero is never imputed for suppression. Primary specifications use a midpoint only when a likelihood or count interval cannot be propagated directly, and every midpoint analysis is accompanied by lower/upper or masking sensitivity.

Scenario envelopes and sharpness. Jurisdiction-level lower/midpoint/upper combinations provide a transparent stress envelope but do not vary every latent cell independently. Cellwise linear optimization can produce an exact coefficient extremum conditional on a fixed design matrix and midpoint exposure. The Wisconsin and California sets are sharp only for their outcome rectangles; they exclude exposure masking, sampling error, other jurisdictions, and common bias. The supplement therefore uses the term sharp only with the conditioning set stated.

Missingness. Nonreporting entities, source-defined not-applicable cells, structural age ineligibility, and suppressed positives are separate statuses. Denominators are audited against category sums and published totals. Rows with impossible category sums or inconsistent eligibility denominators are logged and either excluded or analyzed under a source-specific sensitivity. Coverage dashboards that display no value are not zero-filled.

Measurement error. Classical exposure error attenuates continuous-treatment coefficients toward one, but survey error need not be classical if provider response or exemption reporting correlates with the outcome system. Reliability corrections are therefore sensitivity calculations, not corrected truths. Exposure estimates with published standard errors enter joint bootstrap or combined-variance power calculations when possible.

Ecological mapping. An ecological RR for a ten-point coverage contrast is not the vaccinated versus unvaccinated individual RR. The supplement reports a homogeneous mixture mapping only as an assumption-indexed calibration: the aggregate contrast is diluted by the uptake difference, cohort share in a broad outcome stock, migration, imperfect receipt, and outcome misclassification. Heterogeneous effects, susceptible subgroups, and local treatment responses are not identified. Reduced-form power and mechanically translated individual-scale power are always shown separately.

S33. Multiplicity, specification search, power, Bayesian updating, and synthesis

Multiplicity families. Hypotheses are grouped by data spine, outcome family, exposure family, and timing family before correction. Holm controls familywise error without assuming independence; Benjamini-Hochberg controls the false-discovery rate under its stated dependence conditions. The implemented common-cluster recentered stepdown is labeled an approximate Romano-Wolf-style diagnostic rather than guaranteed finite-sample familywise control because it is not a clearly null-imposed score bootstrap. A result cannot enter a causal pool merely because it survives multiplicity if it fails timing or identification.

Specification curves. The specification universe is defined by defensible choices: link, lag, trend, estimator, exposure definition, outcome denominator, and masking rule. Impossible future lags and outcome-selected ages are excluded. Curves report the full coefficient distribution, but inference is calibrated with common cluster-level wild weights so dependent specifications are not treated as separate studies. Median coefficients and shares significant are descriptive; the maximum positive statistic and simultaneous intervals are the familywise inferential outputs.

Power and equivalence. Each primary result reports its observed contrast, model-conditional standard error, 80 percent MDE, and power at specified moderate effects. Power is calculated on the native reduced-form scale. If a ten-point coefficient extrapolates beyond observed residual variation, the one-within-SD MDE is also reported.

Equivalence requires specified margins and inference appropriate to the native estimator. A common standard-error reconstruction and arbitrary degrees of freedom cannot supply valid TOST p-values across heterogeneous designs. File 430 therefore reports comparisons with published intervals, not reconstructed TOST p-values. Containment of a valid two-sided 95 percent interval within both margins is a conservative sufficient equivalence check conditional on the model; failure of containment is inconclusive. Neither an MDE nor failure to reject a harmful alternative establishes equivalence.

Bayesian calculations. Bayes factors use the approximate normal likelihood and explicit positive half-normal, half-Cauchy, or uniform alternatives. Continuous posteriors use stated prior SDs and material thresholds. Spike-and-slab directional models place visible prior mass on null, harmful, and protective components. Bayes factors do not repair failed counterfactuals, and specification-level posteriors are not multiplied. The local-MMR hierarchy is the most internally commensurable statistical family, although endogenous ecological exposure and suppression uncertainty remain. Section S39 reports an exploratory, assumption-indexed cross-design model only after redefining a coarser scale-dependent policy-contrast parameter; it is not a causal meta-analysis.

Synthesis. Native-scale summaries remain within commensurable estimand families after overlap is removed or dependence is modeled. California state-native and federal IDEA pipelines are alternatives, not replications. Every input to Section S39 is marked `omnibus_pool_eligible = False`; that section deliberately relaxes literal commensurability to show the behavior of a synthetic typical-policy-contrast model. It is reported alongside, never in place of, the family-specific estimates and carries no causal weight. International estimates with incompatible outcomes, failed timing, unmatched ages, or no identified policy contrast still receive zero weight. Modified Hartung-Knapp intervals cannot be narrower than the classical random-effects interval. Small-k funnel tests and confirmatory meta-regression are not used.

All-results accounting. The 25,559 statistical rows include alternative outcomes, lags, masking scenarios, future placebos, failed designs, institutional-shock sensitivities, prenatal screens, and overlapping pipelines. They provide complete row disposition for the frozen release. Among 24,243 auditable ratio-likelihood rows, 1,487 have a directional harm-model posterior above 50 percent and 677 have model-averaged probability above 50 percent for RR exceeding 1.05. Those rows are dominated by overlapping specifications, failed designs, broad outcomes, and zero-weight diagnostics; raw counts of harmful or null posterior screens are not votes. The evidence register records the source layer, native scale, design status, dependence, and zero omnibus weight for every row.

S34. Method-choice robustness matrix and remaining vulnerabilities

Link choice. The flagship log-link RR is 1.013; the identity-link rate difference is +2.6 per 10,000, approximately RR 1.05 at the preperiod mean. Both span no effect. This is modest scale sensitivity, not exact invariance.

Trend choice. Flexible trends eliminate or reverse several national step estimates, which is why England, Spain, Northern Ireland level effects, Sweden, and the national varicella step are not primary causal results. The overlap-safe kink, measured PCV rollout, and local-MMR gradient families retain near-null locations across their defensible trend ladders, with wider uncertainty under exact small-k and common-shock sensitivity.

Estimator choice. Across the six-panel estimator matrix, isolated nominal positive cells do not replicate within their panels. None of the numerically estimable but causally ineligible exemption-event synthetic-control or synthetic-DiD diagnostics rejects under exchangeability-dependent placebo inference. The matrix is not counted as 55 independent tests; it shows whether a substantive result depends on one algorithm.

Inference choice. HC1 can produce smaller p-values in serial panels. State or unit clustering, two-way clustering, wild inference, permutation, spatial covariance, and exact sign flipping generally widen intervals. The paper uses the independent assignment dimension and treats narrower covariance estimators as sensitivities. Local-MMR effective information is approximately 2.11 inverse-variance-equivalent designs despite nine jurisdiction rows.

Exposure choice. PCV, MMR, varicella, and broader schedule uptake are correlated across places. Placebo-exposure estimates reproduce the flagship magnitude, so the causal language is coverage-environment rather than product-pure. RotaShield and Yokohama supply more product-specific first stages but are respectively weak and geographically mismatched. Interaction models require independent coverage support and are zero-weighted when variance inflation or future placebos fail.

Outcome choice. Education-service autism, clinician-reviewed surveillance, specialist-care F84, parent report, broad disability, and functioning yield different estimands. Cross-system agreement is informative when directions and timing align, but these outcomes are never merged as if they measured one latent variable without error. Comparator conditions test specificity only under their own causal exclusion assumptions.

Masking choice. Uniform lower/midpoint/upper scenarios leave the local-MMR point estimate below one, but conditional cellwise sets are wide. The appropriate conclusion is that conventional midpoint findings are not driven by one coherent endpoint choice, while adversarial latent-cell allocation limits sharp population claims.

Pooling choice. Local summaries are separated by denominator and exposure (S67). The historical nine-design leave-one-jurisdiction ratios of 0.991-0.996 and Wisconsin's 65.1 percent weight describe a heterogeneous diagnostic, not an overlap-safe common MMR population effect. PCV kink and local coverage families also remain separate. The S39 cross-design hierarchy's deliberately coarser typical-policy-contrast parameter is not a per-dose or vaccinated-versus-unvaccinated causal RR.

Residual vulnerabilities. Ecological exposure, administrative outcomes, residual cohort trends, migration, diagnosis changes, weak first stages, short preperiods, and inaccessible linked registries remain. Rare product-specific acute adverse events and susceptible-subgroup effects require person-level risk-window designs. The methods above can expose or bound some failures; they cannot substitute for absent treatment variation, clinical outcomes, or controlled linked data.

S35. Reproducibility and computational checks

Source retention. Every acquired official file is retained with URL, access date, release year, raw filename, and hash where redistribution is permitted. Dashboard and API extracts retain query parameters, table identifiers, suppression tokens, and vintage metadata. Failed routes record HTTP behavior and the exact binding block rather than inferring that a field is zero or nonexistent.

Deterministic analysis. Analysis scripts set random seeds for bootstrap, permutation, synthetic-control placebo, and Bayesian simulation steps. Output tables are machine-readable and figures read live values from those outputs. Cohort mapping, transition rules, and design gates are stored as data rather than only prose. The comprehensive validator checks required files, expected dimensions, key numerical reproductions, manuscript phrases, PDF rendering, passive-report exclusions, and absence of unresolved placeholders.

Numerical validation. The residual-method checks pass eleven internal tests; the all-results accounting passes twenty-nine; the comprehensive build passes 101. Midpoint reconstructions for Wisconsin and California match the published equations to numerical tolerance. Fixed positive weights sum to one, dependence standard errors increase monotonically with rho, and all retained dependence weights are nonnegative.

Visual validation. The PDFs are rasterized and inspected page by page. The evidence register is rendered sheet by sheet, with formula-error scanning and source-row reconciliation. Visual review is not a statistical validation, but it prevents clipped labels, stale captions, broken tables, and mismatches between machine-readable results and the paper.

Reproduction boundary. Public-data reproduction covers the archived inputs and scripts. Controlled-data proposals, custodian requests, and future releases are prospective and contribute no estimate. The source inventory is date-bounded and does not certify that every possible dataset has been found.

S36. Objection-to-test register and certainty stress tests

Prebunking protocol. File 2078 registers fourteen recurring objections without conditioning on whether a result is harmful, protective, or null. Each objection is converted into an empirical implication, linked to every compatible place or design, and classified as answered, partly answered, confirmed with narrower causal language, not empirically eliminable from aggregate data, or genuinely outside the data's resolution. The table below is therefore not a list of assurances. It records several limitations that remain binding: product purity is weakened by correlated uptake; Wisconsin dominates inverse-variance precision; arbitrary cellwise suppression allocations are wide; ecological person-level translation is assumption-indexed; and rare susceptible subgroups and acute risk windows require linked individual histories.

Cross-place replication tests. The six-panel estimator matrix contains fifty-five unique panel-by-estimator-family cells. Massachusetts BJS imputation and Czech Poisson fixed effects are the only harmful-direction cells with two-sided p below 0.05. Holm correction leaves zero, and no panel contains two distinct harmful-significant estimator families. The replication rule is deliberately within-panel as well as across-place: a coefficient that appears only under one algorithm is not counted as confirmation. The panel cells remain dependent robustness specifications, so their p-values are not combined as if they came from independent studies. Separately, state-specific PCV7 composition kinks are positive in twenty-three of fifty-one states and nonpositive in twenty-eight. The exact one-sided sign test for an excess of harmful directions is p = 0.799; the median state kink is RR 0.992. Shared national cohort shocks keep this directional sign diagnostic at zero additional causal weight.

Bayesian threshold calibration. File 2081 reports native-scale posterior probabilities under a common symmetric normal prior with SD log(1.10). The probability of RR above 1.05 is effectively zero for the local-MMR family, 1.05 percent for PCV7 rollout, 2.13 percent for RotaShield, and 0.016 percent for the overlap-safe PCV7 kink (a Japan-dominated pool; Japan supplies 53.9 percent of its weight). At the more demanding RR 1.02 threshold, the corresponding probabilities are approximately zero, 31.9 percent, 18.4 percent, and 2.31 percent. These are four separate family likelihoods, not four independent votes and not an omnibus posterior.

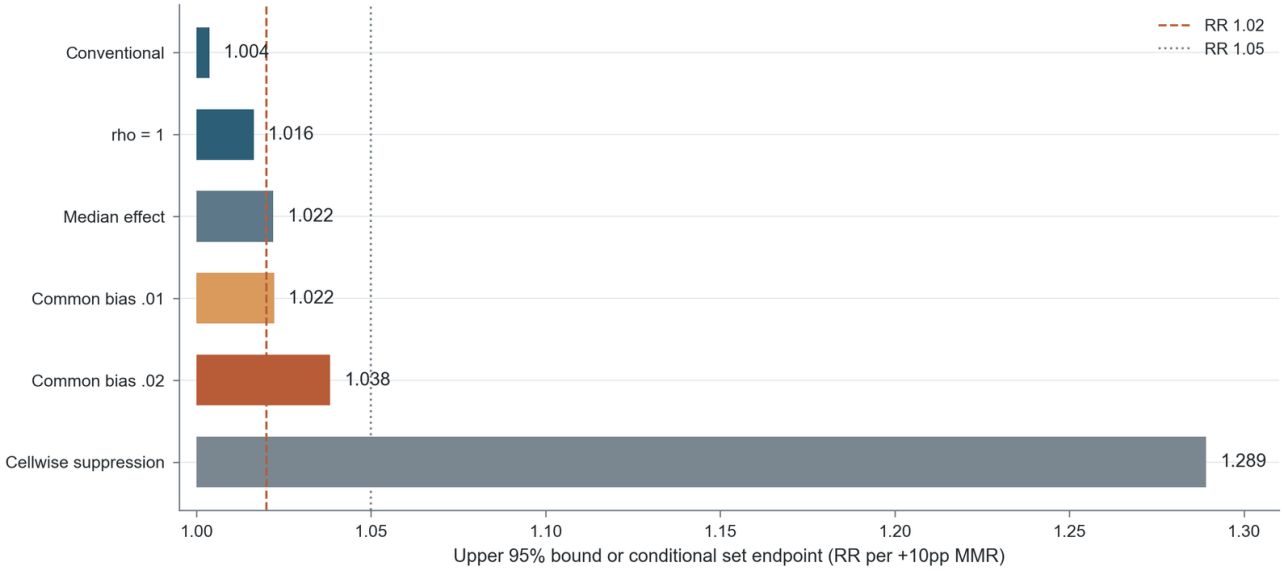


Figure S37. Assumption-indexed certainty stress ladder for the local-MMR family. Conventional, dependence, median-effect, and common-bias rows summarize related but nonidentical targets. The cellwise-suppression endpoint is a conditional partial-identification result, not a pool-wide confidence bound.

Table S13. Objection-to-test register

ID	Objection	Empirical test	Finding	Disposition
A01	The apparent event is a pre-existing cohort trend.	Joint leads, simultaneous bands, one-step-up trends, preperiod-only detrending, splines, and HonestDiD where support permits.	The U.S. measured-rollout exact-age family has two estimated leads and no joint rejection; its longer historical diagnostic crosses an outcome-source transition. Several national steps fail or become curvature-only and are assigned zero primary weight.	answered by downgrade
A02	A nominal policy year misassigns partially exposed cohorts.	Legal start, operational start, dose age, catch-up rule, first fully exposed cohort, transition exclusions, and timing shifts.	Sharp comparisons exclude transition cohorts; blurred varicella and national schedule steps remain diagnostic rather than RD estimates.	answered with scope limit
A03	Serial or spatial dependence makes conventional p-values too small.	State/unit and two-way clustering, Driscoll-Kraay, Conley, wild cluster, exact 2^9 sign flipping, and Wisconsin county clustering.	Dependence widens intervals but creates no robust harmful result; local-MMR rho=1 gives upper RR 1.016 and Wisconsin county clustering gives upper RR 1.002.	answered for focal families
A04	The result is one estimator's artifact.	TWFE, first/long differences, Poisson, imputation DiD, Callaway-Sant'Anna, Sun-Abraham, stacked DiD, synthetic control, SDID, and matrix completion where eligible.	Among 55 unique estimator-by-panel cells, two are nominally harmful at p<.05, zero survive Holm correction, and no panel has two distinct harmful-significant estimator families.	new cross place test answered
A05	One state or jurisdiction drives the pooled null.	Leave-one-state/jurisdiction/donor-out checks, influence functions, and positive-weight effective-design counts.	Local-MMR leave-one-out RRs are 0.991-0.996, although Wisconsin supplies 65.1% of inverse-variance weight and the effective design count is 2.11.	partially answered dominance remains
A06	Correlated uptake prevents product-specific attribution.	Placebo-exposure substitution, multivaccine models, covariance/support audits, and variance-inflation checks.	MMR and varicella placebo exposures reproduce the PCV7 magnitude; pair interactions do not survive timing and multiplicity gates.	objection confirmed language narrowed
A07	Outcome ascertainment or category substitution creates the pattern.	Independent surveillance/education/survey systems, autism-plus-speech, autism share, DDD, and joint multivariate wild inference.	No coherent autism-specific harmful pattern survives; comparator outcomes move heterogeneously and are treated as claim-relevant outcomes, not assumed negative controls.	partially answered outcomes differ
A08	Suppression and masking hide a harmful association.	243 coherent endpoint scenarios and conditional cellwise linear optimization for dominant Wisconsin and California equations.	Coherent scenarios stay near RR 0.991-0.993, but conditional cellwise sets widen to 0.801-1.212 in Wisconsin, 0.520-1.871 in California, and 0.757-1.289 for the dominant-two fixed-weight pool.	partially answered adversarial bounds wide
A09	Estimated vaccine coverage is measured with sampling error.	Kish-Beta redraws, combined outcome/exposure variance, reliability stress, and joint NIS-error propagation.	Combined upper RRs are 1.046 for PCV7 and 1.051 for RotaShield; shared NIS exposure-error correlations are at most 0.13.	answered for measured us exposures
A10	Specification search manufactures significance.	Specified specification universe, common ild weights, max-t FWER, Holm/BH/Romano-Wolf, and complete result accounting.	The 56-specification state-MMR curve has familywise p=0.685 for its largest positive t and zero simultaneous intervals excluding one; 0/55 cross-place estimator cells survive Holm for harm.	answered for registered families
A11	A common reporting shock makes place estimates look independent.	Fixed-positive-weight rho sensitivity and hierarchical additive common-shock SD grid.	At rho=1 the upper RR is 1.016. A 0.02 log-RR common-shock SD raises the hierarchical upper interval to about 1.038 and P(mean RR>1.02) to 11.9% under the widest retained prior grid.	partially answered bias not identified
A12	Ecological coefficients cannot identify an individual vaccinat ed-versus-unvaccinated RR.	Homogeneous mixture mapping, dilution tables, individual observational corroboration, and explicit separate estimands.	The translation is reported only as assumption-indexed calibration; it is not pooled with person-level studies.	not empirically eliminable with public aggregate data
A13	Rare susceptible subgroups or acute risk windows are averaged away.	Scope analysis and comparison with known positive-control adverse-event designs.	The held public panels cannot resolve these effects; controlled linked registers are required.	genuinely not testable here
A14	Placebo failure is ignored or repaired by adding controls.	Future exposure, younger unaffected cohorts, fake cutoffs, in-time placebos, and non-event cutoff ranks.	Peru and national varicella fail timing diagnostics and receive zero causal weight; NJ, Oregon, Arizona, and Ohio future terms span no effect.	answered by symmetric gate

Notes: Partly answered and not testable denote limits of the available evidence. File 2078 records empirical implications, settings, certainty effects, and source provenance.

Interpreting the stress ladder. The first five rows progressively relax sampling and common-shock assumptions: conventional t8 inference, perfect positive sampling correlation with fixed weights, a distribution-free interval for the median jurisdiction, and common-bias SDs of 0.01 and 0.02 log-RR. Their upper endpoints are 1.004, 1.016, 1.022, 1.022, and 1.038 per ten MMR coverage points. Under the widest retained prior grid, the 0.02 common-bias sensitivity leaves 11.9 percent posterior probability that the mean RR exceeds 1.02 but only 0.76 percent above 1.05. The last row changes the object: cellwise optimization of dominant Wisconsin and California outcome rectangles reaches 1.289 while fixing midpoint exposure, positive pooling weights, and the other seven equations. It is a conditional partial-identification warning, not a sampling interval or a pool-wide set.

What no robustness test can supply. Flexible models cannot manufacture untreated cohorts for a national step, recover sibling-specific histories absent from public files, infer a product-specific effect from collinear coverage, or

identify an acute susceptible-subgroup response from annual aggregate stocks. The supplement therefore uses failed placebos and failed pretrends to downgrade designs, and labels untestable objections rather than claiming that more controls resolved them. The strongest defensible conclusion concerns generalized population effects on the covered ecological and cohort scales; certainty is necessarily lower for very small effects, rare subgroups, acute windows, and products without a sharp public first stage.

S37. Institutional explanations for recorded diagnosis

S37.1 Rival theories and observable implications

The theoretical audit separates two data-generating processes. Under a generalized chronic vaccine-harm theory, schedule additions or higher uptake increase underlying disorder risk. The observable implications are cohort-timed increases after additions, decreases after withdrawals, monotone uptake gradients, larger effects where the exposure contrast is larger, incremental effects when a vaccine is added to an existing schedule, and recurrence across independent outcome systems. Under a diagnosis-production theory, underlying traits can be stable while recorded prevalence changes because classification, screening, referral, service eligibility, reimbursement, provider supply, or reporting changed. Its implications are calendar- or policy-timed breaks, changing age at diagnosis, category substitution, sensitivity to outcome source and service capacity, and continued administrative growth after exposure saturation. These institutional mechanisms are genuine causes of recorded prevalence, not nuisance variation to be removed; the identification problem is whether their timing and geography can be separated from vaccine exposure.

The theories are not exhaustive and need not be mutually exclusive. The purpose of the register is to ask which distinctive implications are supported. It does not infer that diagnosis production explains all autism growth, and it does not declare a vaccine theory false merely because administrative outcomes are imperfect. File 2086 records twelve institutional mechanisms, their predictions, the implemented tests or design protections, results, and residual limitations. File 2088 records the supporting statutes, official guidance, and peer-reviewed literature.

S37.2 Institutional timeline and why it matters

IDEA classification. The October 1990 amendments added autism and traumatic brain injury as explicit disability categories. The resulting administrative autism series begins in 1991-92; its early growth cannot be read as incidence growth without accounting for category adoption. Child Find requires states and LEAs to identify, locate, and evaluate children suspected of disability. IDEA classification determines eligibility and reporting, not a uniform clinical diagnostic process.

IDEA finance. Before the formula transition, federal Part B grants depended on the number of children reported as served. The 1997 amendments established an FY1999 base and distributed incremental funds 85 percent by eligible-age population and 15 percent by child poverty. The Department of Education states that the change sought to remove inappropriate-identification incentives. The transition was complete before 2000, while the flagship age-eight outcomes are observed in 2005-2011. It is therefore outside the flagship estimation window. This removes a direct coincident national break, but it does not measure state and local formulas or heterogeneous persistent effects. Cullen (2003) and Dhuey and Lipscomb (2011) show empirically that fiscal incentives and capitation reforms can change administrative disability rates.

Clinical identification and reporting. AAP guidance in 2007 recommended autism-specific screening at 18 and 24 months. The flagship birth cohorts are 1997-2003, so none was age 18-24 months when that guidance appeared. Common later ascertainment shocks are absorbed by cohort effects, but heterogeneous state or provider adoption is unobserved. DSM-5 appeared in 2013, after the flagship outcome window. Earlier DSM/ICD changes remain relevant to historical and international series. Parner, Schendel, and Thorsen (2008) show that falling age at diagnosis can inflate young-cohort comparisons, and Hansen, Schendel, and Parner (2015) estimate that Denmark's 1994 criteria change and 1995 outpatient inclusion explain 60 percent of the observed increase in the studied cohorts.

Service incentives. Insurance mandates began in Indiana in 2001 and expanded in waves. Mandell et al. (2016) estimate a 0.21-per-1,000 increase in commercially treated ASD prevalence, 12.7 percent relative to the no-mandate baseline, with a larger effect after several years. Medicaid state-plan, EPSDT, and HCBS pathways can similarly make diagnosis a gateway to services. LaClair et al. (2019) identify nine states with child-specific autism HCBS waivers during 2003-2011 and link waiver generosity to unmet-need disparities; Shea et al. (2021) document 269 Medicaid programs in a fifty-state 2004-2015 review and a more-than-fivefold increase in autism-specific 1915(c) waivers. These

mechanisms predict treated or administrative prevalence, not biological incidence. Mandate dates and first-observed child-specific waiver years overlap the flagship and can be tested directly; a complete state-year series of exact enactment dates, generosity, enrollment caps, waiting lists, enforcement, provider capacity, and district funding rules is not held and is not claimed to be controlled.

S37.3 Direct mandate-sensitive flagship models

For state s and birth cohort c, the preferred conditional mean is

$$\mathbb{E}[Y_{sc} \mid X] = \exp\left(\alpha_s + \lambda_c + \beta R_s \mathbb{I}_{\{c \geq 2000\}} + R_s g(c) + \delta P_{s,c+8} + \log N_{sc}\right)$$

where Y is IDEA autism at age eight, N is the age-eight resident population, R is standardized early PCV7 uptake, g(c) is the low-rank cohort spline interacted with uptake intensity, and P is the institutional policy term. State-clustered standard errors use 51 clusters and t critical values with 50 degrees of freedom. The policy sensitivities replace P with a mandate indicator, mandate-duration bins, child-specific HCBS-waiver status, joint mandate-and-waiver controls, observation censoring, adoption-window exclusion, or state-specific linear cohort paths. They preserve the treatment definition and population offset. The waiver variable begins in the first year a state is observed with a waiver in the LaClair et al. inventory; it is not represented as an exact legal-effective date.

Institutional sensitivity	RR	95% CI	Rows
State and cohort fixed effects	1.017	0.979-1.057	353
Preferred flexible intensity path	1.013	0.981-1.045	353
Plus insurance-mandate indicator	1.019	0.985-1.054	353
Plus mandate-duration bins	1.018	0.985-1.052	353
Plus child-specific HCBS-waiver indicator	1.013	0.981-1.046	353
Plus mandate and HCBS-waiver indicators	1.019	0.985-1.054	353
Censor every post-HCBS-waiver observation	1.010	0.975-1.046	315
Censor every post-mandate observation	1.009	0.976-1.044	295
Exclude year -1 through year +3 around adoption	1.019	0.980-1.059	274
Exclude latest birth cohort	1.017	0.985-1.051	302
State-specific linear cohort paths	1.014	0.980-1.050	353

The point-estimate range is 1.009-1.019 and the largest absolute movement from the preferred RR is 0.006. Every interval crosses one; every upper endpoint is below 1.060. The diagnosis-policy coefficients themselves are reported in file 2091. In the flagship, the post-mandate coefficient is RR 1.056 (0.999-1.117; p = 0.054); duration-bin RRs are 1.044 in year one, 1.054 in year two, and 1.088 after three or more years. The child-specific HCBS-waiver coefficient is RR 1.017 (0.937-1.104; p = 0.673). These associations are not interpreted causally because adoption can respond to latent service demand and waiver timing is interval-censored by the published inventory. In the separate SchoolVaxView-IDEA composition panel, adding the mandate indicator changes the MMR coefficient from RR 1.065 (0.964-1.176) to 1.067 (0.968-1.175). The post-mandate coefficient in that education outcome is 1.010 (0.955-1.069), but that is not a failed replication of Mandell et al.: their estimand is privately treated prevalence among mandate-eligible children, whereas this outcome is the autism share of IDEA enrollment. The robustness question is whether controlling the measured diagnosis-service shock changes the vaccine coefficient. It does not materially do so.

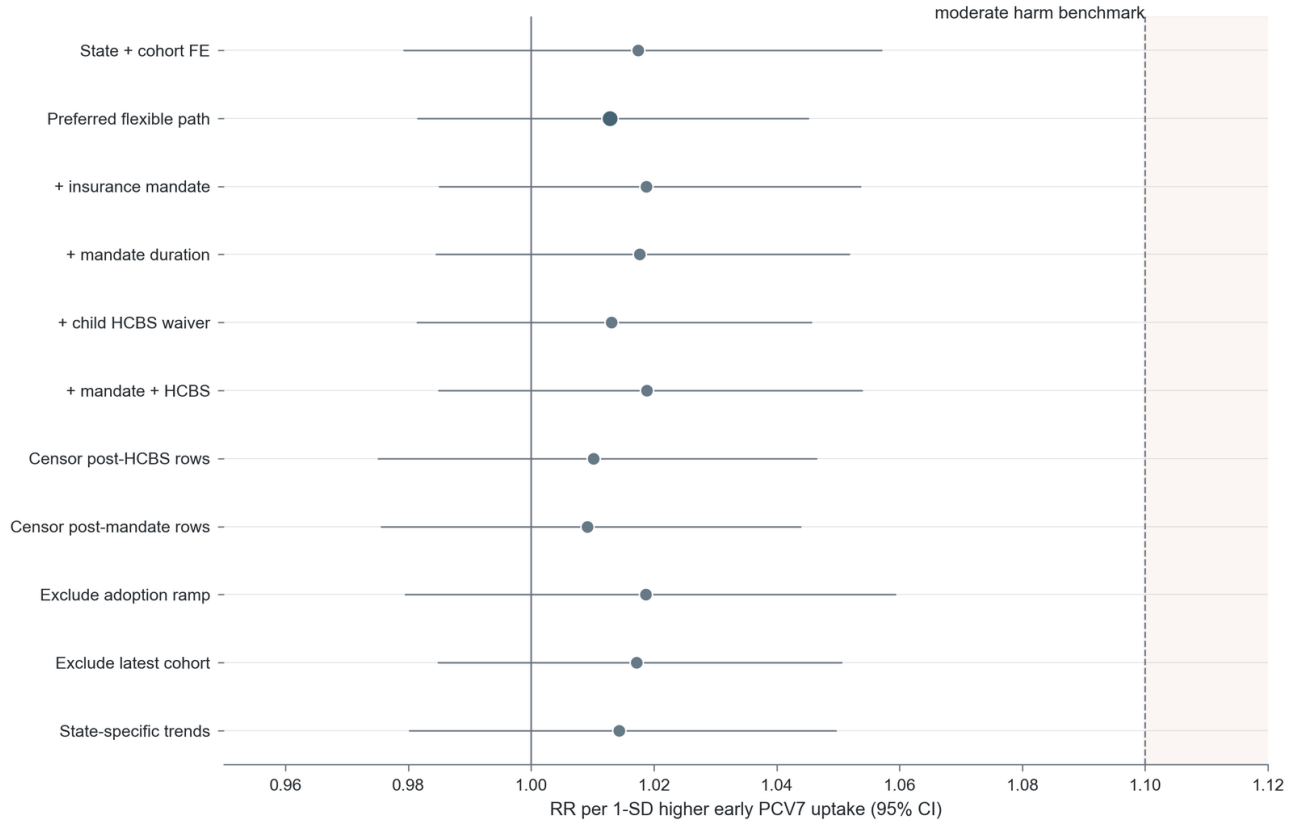


Figure S38. Institutional diagnosis-incentive robustness for the flagship U.S. PCV7 cohort design. Insurance-mandate and child-specific HCBS-waiver controls, joint policy adjustment, policy-exposed observation censoring, mandate-duration bins, adoption-window exclusion, endpoint deletion, and state-specific cohort paths leave the exposure coefficient near one. Waiver dates are first-observed inventory years, not exact enactment dates; the figure does not claim to observe every state or local funding rule.

S37.3a Exact-age saturation for heterogeneous institutional shocks

National policy shocks are absorbed by cohort effects only if their outcome effect is common across states. To relax that assumption, file 2092 reconstructs autism counts at each exact age six through eleven for 2005-2011, joins exact-age Census denominators, and estimates

$$\log\left(\frac{Y_{say}+0.5}{N_{say}}\right)=\beta R_s\mathbb{I}_{\{y-a\geq 2000\}}+\alpha_{sy}+\gamma_{sa}+\lambda_{ay}+\varepsilon_{say}$$

where s is state, a is exact age, y is report year, R is standardized early PCV7 uptake, alpha_sy is a complete state-by-report-year effect, gamma_sa is a state-by-age effect, and lambda_ay is an age-by-report-year effect. The alpha_sy terms absorb arbitrary state-specific annual changes in IDEA funding, Child Find, insurance mandates, service capacity, reporting, or diagnostic intensity that shift the observed ages together. The other two fixed-effect sets absorb persistent state age profiles and national age-specific reporting shocks. Population-weighted least squares uses state-clustered inference with 50 degrees of freedom. A rank-revealing QR decomposition retains a full-rank nuisance basis because the three dummy families are algebraically redundant.

Exact-age institutional-shock model	RR	95% CI	p	Cells
Population weighted	1.006	0.967-1.047	0.749	2,108
Unweighted	0.993	0.962-1.025	0.670	2,108

This is an exposed-cohort step, not a free event-study path. With state-by-year, state-by-age, and age-by-year fixed effects, unrestricted cohort-specific intensity coefficients encounter the affine age-period-cohort identity; identifying a dynamic path would require an additional normalization. The flagship event-study leads therefore provide the pretrend evidence. The saturated model also cannot absorb an unmeasured state-by-year-by-age shock that lines up exactly with

PCV-eligible cohorts. It materially weakens the objection that a heterogeneous state-level institutional expansion alone drives the preferred estimate, but it does not make that exclusion restriction assumption-free.

S37.4 Funding, category substitution, and the outcome vector

The federal formula change is not presented as directly falsified. Its national timing is outside the flagship window, and common shocks are absorbed by cohort effects under a uniform-effect assumption. The residual threat is a state-specific time-varying funding or implementation change correlated with early PCV7 uptake. Five features probe that threat without pretending to observe every policy: arbitrary state-specific linear cohort paths; the flexible intensity-specific counterfactual; the exact-age state-by-year saturated model; autism as a share of total IDEA enrollment; and the joint disability-category vector.

Per ten MMR coverage points in the SchoolVaxView composition panel, the category RRs are 0.901 for specific learning disability, 0.951 developmental delay, 0.968 traumatic brain injury, 1.001 emotional disturbance, 1.027 intellectual disability, 1.050 speech/language, 1.063 autism, and 1.168 (95% CI 1.013–1.347; p = .032, without multiplicity adjustment) other health impairment. Autism is neither the largest category association nor statistically distinct from one. Other health impairment is nominally positive, while the external-cause benchmark of traumatic brain injury is null. The preferred PCV7 diagnostic-substitution DDD is likewise near one, and the complete multivariate wild-bootstrap category family produces no coherent autism-specific harmful result.

Diagnostic substitution could nevertheless hide a true increase if autism rises while intellectual-disability or speech labels fall. File 2093 therefore sums the administrative counts before estimation under the same preferred flexible PCV7 model. Autism plus intellectual disability plus speech/language gives RR 1.000 (0.964-1.036; p = 0.980); autism plus speech gives 0.997 (0.961-1.035); and autism plus intellectual disability gives 1.018 (0.983-1.054). These are administrative category pools rather than clinical composites, but they directly test the arithmetic masking mechanism.

This test is not described as eight negative controls. A broad vaccine-injury claim can encompass autism, other neurodevelopmental conditions, speech/language, intellectual disability, psychiatric conditions, epilepsy, physical impairment, hearing, or vision. Those are claim-relevant outcomes. Their joint pattern asks whether a proposed generalized injury mechanism is coherent and specific. Traumatic brain injury has a stronger external cause and therefore serves as a benchmark. Future exposure and fake dates are temporal controls. Vaccine-preventable infections are positive-control disease families. The causal role of each outcome is justified separately in file 2087.

S37.5 Institutional objections and empirical conclusions

Classification and funding mechanisms:

- IDEA autism-category creation is directly visible and accepted as a cause of the early administrative series break. It motivates stable-era, fixed-age analysis and category-combination tests; it is not a vaccine effect.
- The federal count-based incentive and its 1997 reform are protected by timing and cohort effects in the flagship but not globally falsified. State and local formula histories remain incomplete.
- Child Find and educational eligibility are partly tested through category composition, total-IDEA denominators, and cross-system replication. No complete enforcement index is observed.
- State special-education formulas, district fiscal constraint, and school resources are partly tested through state-specific paths and smaller-area replication. A harmonized annual policy panel remains absent.

Diagnostic and reporting mechanisms:

- DSM/ICD and reporting changes are documented causes of source-specific prevalence breaks. DSM-5 postdates the flagship, while earlier and international changes are handled by stable windows and source-specific trends where possible.
- AAP screening and falling age at diagnosis are mitigated by a fixed age-eight endpoint and equal follow-up. State/provider adoption remains unmeasured.
- Diagnostic substitution and accretion are accepted historical mechanisms and directly tested with neighboring categories and the DDD. They do not produce a vaccine-timed harmful differential.
- A single outcome-pipeline artifact is probed with IDEA, ADDM, NSCH, NHIS, England SEN, Swedish specialist care, and Brazil Census. Related state and federal publications sharing the same administrative spine are counted once.

Access and composition mechanisms:

- Insurance mandates are directly controlled, duration-modeled, censored, and window-excluded; the vaccine coefficient is stable.
- Child-specific autism HCBS waivers are directly controlled, jointly modeled with mandates, and post-waiver observations censored. The corresponding vaccine RRs are 1.013, 1.019, and 1.010. The direct waiver coefficient is imprecise and the first-observed timing cannot encode exact effective dates, generosity, caps, waiting lists, or provider capacity.
- Medicaid and EPSDT pathways are source-stratified and labeled treated prevalence. The 2014 federal autism guidance postdates the flagship. The held Medicaid panel demonstrates the danger directly: its naive PCV RR is 1.336 (1.115-1.602), yet an impossible future schedule share is even larger at 1.453 (1.119-1.885), all fourteen naive fake cutoff years reject, and MAX- versus TAF-era estimates disagree. This is a reporting/trend failure, not vaccine evidence, and it shows why administrative policy systems require temporal diagnostics rather than nominal-significance counting.
- Public awareness, evaluator supply, and service access remain plausible causes of secular growth. Saturation and cross-system tests show that measured vaccination drift cannot generate that growth, but they do not identify which institutional factor did.
- Migration and family sorting are partly addressed with resident denominators, fixed geography, and local panels. Only linked person-level residence and vaccination histories can remove exposure-geography misclassification completely.

S37.6 What the institutional evidence changes - and what it does not

The evidence supports a narrow conclusion. The flagship estimate is not materially altered by the time-varying diagnosis incentives observed with usable state timing, including insurance mandates and child-specific HCBS waivers, and its result is compatible with category, trend, and independent-outcome-system checks. These analyses do not attempt to subtract policy-driven diagnosis from history. They ask whether genuine institutional changes are aligned with the vaccine contrast strongly enough to be misattributed to vaccination. The evidence also explains why raw administrative autism counts can rise rapidly without a corresponding change in underlying incidence. Funding, screening, criteria, service access, and reporting must be treated as substantive rival mechanisms.

It does not make the analytical choices unimpeachable. State funding reforms are incompletely harmonized; insurance laws vary in ERISA reach, age limits, dollar caps, and enforcement; Medicaid waiver capacity and waiting lists are not fully observed; provider and school-resource panels are incomplete; and migration remains ecological. Those omissions are not repaired with more fixed effects. They are stated as residual identification threats and motivate controlled linked-register work. The evidence still strongly constrains a large generalized, coverage-proportional, cohort-timed vaccine effect; it does not identify or exclude every very small, rare-subgroup, acute-window, or policy-mediated individual effect.

S38. Substantive causal theories beyond diagnosis production

S38.1 Theory-to-prediction discipline

Section S37 explains how recorded disability can change without biological incidence changing. This section evaluates a different class of objections: proposed biological, behavioral, and selection pathways through which vaccination might change true outcomes. File 2094 contains the twenty-two-theory register; file 2095 links each theory to twenty-four available or explicitly missing evidence modules; file 2096 records sixteen primary literature sources; and files 2097-2100 contain the prenatal screen, subgroup summary, validation, and machine-readable results. Empirical implications are specified independently of coefficient sign. Evidence roles remain separate: quasi-experimental estimates, falsifications, descriptive screens, positive-control checks, mechanism-only evidence, and adjusted individual comparisons are not pooled merely because they address the same verbal theory. Individual observational estimates form a separate triangulating layer, receive zero policy-design weight, and never narrow the policy-design confidence or posterior intervals.

ID	Proposed theory	Observable implication	Strongest held test	Conclusion	Residual limitation
T01	Product or antigen specificity	The named product, not correlated schedule coverage, predicts cohort-timed change and reverses on removal	PCV7, MMR, RotaShield, thimerosal and withdrawal families	No product-specific harmful pattern clears timing, trend, placebo, and multiplicity gates	Uptake is correlated; public files lack brand and lot identifiers
T02	Persistent measles infection or enteropathy	Vaccine-strain RNA and ordered GI-to-developmental onset after MMR	Blinded bowel-tissue replication and MMR timing/withdrawal evidence	Specific virological and temporal predictions were not observed	Does not deny GI comorbidity in autistic people
T03	Thimerosal ethylmercury toxicity	Risk rises with prenatal or infant dose and falls after removal	Observational cohort and case-control calibration; removal histories	Observational calibration reports no dose response; causally weak removal series show no decline	Other mercury sources and rare metabolic strata are different estimands
T04	Aluminum-adjuvant chronic toxicity	Neurodevelopmental, autoimmune, and allergic risk rises with cumulative dose	1.22-million-child observational calibration and broad outcome screens	No increased association across the measured fifty chronic outcomes; zero causal weight	Very rare syndromes and exposure misclassification remain
T05	Cumulative antigen or immune overload	ASD, including regression, rises with early cumulative antigens and schedule size	Observational antigen calibration; NSCH schedule gradients; saturation tests	Observational gradient is null; cohort-aligned schedule and saturation predictions fail	Antigen count is not identical to injection or innate-response burden
T06	Multi-vaccine synergy	Adding B to schedule A raises the outcome incrementally; with factorial support, the A-and-B joint effect also departs positively from additivity	Twenty-one-row interaction set, PCV-by-baseline, England nested states, Oregon pairs	No interaction passes joint-support, timing, trend, and multiplicity gates; nested incremental estimates remain separate	Most public schedules do not supply factorial exposure support
T07	Critical age window or latency	Effects vary systematically with administration age and exact post-dose lag	Cross-event age gradient, NHIS timing, self-controlled literature, lag grids	No monotone age gradient; public timing estimates are sparse	Cross-product age comparisons confound product with age; annual data miss days-long windows
T08	Genetic or medical susceptibility	Larger effects in familial, genetic, preterm, metabolic, or clinical-risk strata	Familial-risk cohorts and NHIS measured subgroup battery	No harmful familial-risk MMR result and no Holm-robust NHIS heterogeneity	Genetic, metabolic, and adequately powered rare clinical strata are absent
T09	Nonlinear threshold or saturation	A breakpoint appears above the threshold and growth levels after stable uptake	Quadratic intensity, schedule-state, and saturation analyses	Simple nonlinearity unsupported; autism identification continues after stabilization	Rare person-level thresholds may be invisible in aggregate coverage
T10	Maternal or prenatal vaccination	Outcomes track pregnancy-specific exposure by birth cohort	Prenatal thimerosal and Tdap cohorts; Brazil maternal-Tdap screen	No increased ASD in the linked Tdap cohort or ecological screen	Maternal influenza exposure and product-specific pregnancy linkage remain
T11	Acute fever, seizure, or regression trigger	Named phenotype clusters within days or weeks of a specific product and dose	Self-controlled MMR literature; MMRV febrile-seizure and RotaShield benchmarks	Real rare harms show product/window specificity; annual stocks cannot adjudicate every acute claim	Validated onset and exact dose dates are required
T12	Infection-mediated net effect	Vaccination reduces infection; infection shocks mediate later disability risk	Brazil hospitalization positive-control families	Expected protection was not recovered, weakening that panel's validation	A failed positive control is neither evidence of vaccine injury nor a complete mediation test
T13	Healthy-vaccinee selection, health-care contact, or reverse causation	Crude individual associations change with adjustment; policy contrasts differ from family-choice contrasts	NHIS adjustment ladder, future exposure, exemption and policy designs	Selection is a confirmed threat; quasi-experiments avoid some but not all ecological confounding	No public sibling-fixed-effect or randomized-encouragement design
T14	Mortality or competing-risk masking	Mortality or severe morbidity rises enough to remove cases; combined outcomes reveal displacement	Autism-ID-speech pools and mortality feasibility audit	Category displacement not supported; autism-coded mortality too sparse for broad bounds	Exposure-linked cause-specific mortality is not held

ID	Proposed theory	Observable implication	Strongest held test	Conclusion	Residual limitation
T15	Broad chronic multisystem injury	Claimed psychiatric, neurological, sensory, allergic, autoimmune, and physical families move coherently	IDEA, Norway, Brazil, France, Taiwan, international families; Danish cohort as calibration	No coherent harmful chronic pattern survives the diagnostic sequence	Several outcomes measure care use or administrative composition, not incidence
T16	Mitochondrial, redox, or neuroimmune vulnerability	Effects concentrate in biomarker-defined children in an acute plausible window	Developmental/genetic literature and familial-risk strata	Autism biology does not identify vaccination as the initiating exposure	No biomarker-defined vaccinated versus unvaccinated contrast
T17	Spacing, catch-up, or simultaneous visits	Equal cumulative dose with different visit spacing gives different risk	NHIS timing/dose counts and transition-cohort checks	No robust public timing signal; simultaneous visits remain weakly observed	Recall, sparse cases, and selection into delay/catch-up
T18	Autoimmune or molecular-mimicry pathway	Product- and latency-specific autoimmune disease and later neurodevelopmental change	Danish autoimmune/allergic families and broad screens	No generalized chronic autoimmune or allergic signal	Rare validated encephalitic syndromes are not equivalent to autism prevalence
T19	Microbiome, barrier, or gut-brain pathway	Product- and time-specific microbiome or barrier change precedes developmental outcome independently of antibiotics, diet, infection, and care contact	No prospective pre/post-dose microbiome panel is held	Not directly tested; aggregate GI patterns cannot identify mediation	Requires pre-exposure multi-omic sampling and validated outcomes
T20	Non-Tdap maternal immune activation	Pregnancy product, trimester, fever, and infection histories predict offspring outcomes in separable ways	Tdap and prenatal ethylmercury only	Product-specific evidence does not generalize to influenza or every maternal exposure	Confounding by indication, infection, fever, trimester, and care contact
T21	Epigenetic or transgenerational pathway	Exposure-linked tissue-relevant epigenetic change precedes and mediates phenotype across generations	No multigenerational biospecimen data are held	Not testable with the project data	Ecological schedule timing is not evidence of molecular mediation
T22	Lot, contaminant, residual excipient, or manufacturing-specific toxicity	Risk clusters by verified brand or lot and a reproducible laboratory assay	No case-linked lot or contaminant assay is held	Not testable with aggregate coverage; no causal lot cluster identified	Requires lot-linked denominator data, clinical validation, and laboratory confirmation

S38.2 Prenatal maternal-Tdap cohort screen

The public-data screen joins SI-PNI maternal-Tdap coverage to the 2022 Brazilian Census autism and disability tables. For municipality m and Census age band a, exposure is the mean coverage in the approximate birth window: 2018-2022 for ages zero to four and 2013-2017 for ages five to nine. A municipality enters only when both age groups are observed and at least four annual coverage values are available within each window. The primary model is

$$\log\left(\frac{Y_{ma}+0.5}{N_{ma}+1}\right)=\beta\frac{V_{ma}}{10}+\alpha_m+\lambda_{\mathrm{state}(m),a}+\varepsilon_{ma}$$

where Y is the reported case count, N is the Census population, V is maternal-Tdap coverage in percentage points, alpha_m is a municipality effect, and lambda_state(m),a is a state-by-age-band effect. Population weights are primary; standard errors are clustered by municipality. The coefficient is a within-municipality contrast between two broad age groups after state-specific age profiles. It is not an individual pregnancy effect, a policy instrument, or a sharp cohort discontinuity.

Five specified exposure treatments address SI-PNI denominator anomalies: annual coverage capped at 100 before taking the mean; annual values above 100 excluded; annual coverage capped at 100 before taking the median; the raw annual mean; and an unweighted outcome regression using the capped mean. Autism is estimated for total, male, and

female populations. Any disability is estimated as a broader claim-relevant outcome. Impossible future or post-birth exposure assigns the later maternal-coverage window to older already-born cohorts. The result file contains thirty models in total.

The primary total-autism estimate is RR 1.000 per ten coverage points (95 percent CI 0.988-1.012; p = 0.943; 6,378 age cells in 3,189 municipalities). Excluding annual coverage above 100 gives RR 0.998 (0.985-1.011); using the capped median gives 0.999 (0.990-1.008); retaining the raw mean gives 1.001 (0.989-1.013). Two unweighted cohort-aligned secondary models are nominally positive: male autism RR 1.019 (1.004-1.035; p = 0.011) and any disability RR 1.019 (1.002-1.035; p = 0.024). Neither pattern is robust to population weighting and the exposure-cleaning variants, but both are retained in file 2097 rather than selected by sign. Sex-specific, any-disability, future-exposure, and all other unweighted results are also retained. Every row has causal synthesis weight zero because one Census, five-year bins, ecological exposure, age-varying diagnosis opportunity, and imperfect administrative denominators do not create exchangeability.

The individual literature measures pregnancy exposure more directly but does not identify a causal effect. Becerra-Culqui et al. (2018) studied 81,993 mother-child pairs and reported adjusted HR 0.85 (0.77-0.95) for offspring ASD after prenatal Tdap; Price et al. (2010) reported no prenatal or infant ethylmercury association. These observational results are retained solely for calibration. They receive zero causal weight, are not meta-analyzed with the Brazil screen, and do not strengthen the paper's causal conclusion.

S38.3 Susceptibility, timing, dose, and interactions

The NHIS subgroup archive contains 376 coefficient rows across autism, ADHD, and learning disability. For autism, 110 rows were attempted and 47 adjusted rate ratios are finite. Fourteen exposure-by-subgroup heterogeneity tests are estimable. Neither an individual autism coefficient nor a heterogeneity test is significant after Holm correction. The strata cover sex, age, survey era, income, race/ethnicity, and maternal education; they do not cover genotype, mitochondrial phenotype, prematurity in adequate numbers, or verified sibling-specific exposure. These sparse public-use associations are calibration of measured strata only, receive zero causal weight, and cannot establish the absence of every susceptible subgroup.

The age-at-exposure audit separates within-product timing from a cross-product meta-regression. The latter estimates a log-RR slope of -0.0022 per month (p = 0.530), corresponding to a twelve-month-versus-two-month RR ratio of 0.978 (0.893-1.071). This does not identify a causal age effect because product and administration age co-vary. NHIS first-dose and dose-count estimates are individual but rely on recalled or public-use timing, sparse autism cases, and selection into delayed or catch-up vaccination. Exact-date self-controlled designs are the appropriate method for a days-long acute mechanism.

Interaction tests require support for all relevant schedule states. The preferred PCV-by-baseline schedule interaction is RR 1.001 (0.989-1.012 after joint NIS exposure uncertainty). The broader twenty-one-row interaction summary contains no pair that simultaneously satisfies exposure support, timing, trend, placebo, and multiplicity gates. High correlation between schedule components is treated as an identification limit, not solved by adding more product-by-product interaction terms. Exact visit histories with natural or policy-induced separation of simultaneous and staggered administration remain the target design.

S38.4 Acute harms, broad conditions, and positive controls

Other conditions are not automatically negative controls. A claim that vaccines cause intellectual disability, ADHD, epilepsy, hearing loss, autoimmune disease, allergy, or physical impairment makes those outcomes targets of the claim. Their role is coherence and specificity: a generalized multisystem theory predicts an organized product-, timing-, and dose-specific pattern, not isolated nominal coefficients scattered across heterogeneous administrative systems. File 2095 therefore labels each analysis as causal, falsification, descriptive, mechanism-only, failed feasibility, or positive-control evidence.

Known rare adverse events provide external method benchmarks. Product-specific risk-window and postlicensure studies found an MMRV febrile-seizure excess in days seven through ten compared with separately administered MMR and varicella, and found acute RotaShield-associated intussusception. The public cohort panels used here do not contain linked dose dates and acute clinical onset dates, so they neither reproduce nor test those days-to-weeks harms. In particular, the RotaShield analysis estimates later autism identification after a low-uptake pulse, not infant

intussusception. These events have named products, phenotypes, and narrow latency windows. Their existence prevents the framework from defining safety by assumption and also shows why annual autism or disability prevalence cannot exclude every acute post-dose event. Conversely, the failure of Brazil hospitalization analyses to recover expected protection against selected vaccine-preventable infections weakens those hospitalization panels; it is not evidence of vaccine injury.

The benchmark studies carry the detail that makes them credible positive evidence. The Vaccine Safety Datalink compared 83,107 first-dose MMRV recipients with 376,354 recipients of separate same-day MMR plus varicella and found RR 1.98 (1.43-2.73) for seizures during days seven through ten, an excess of 4.3 per 10,000 doses, or about one additional febrile seizure per 2,300 MMRV doses (Klein et al. 2010). Nationwide surveillance during the 1976 A/New Jersey influenza program found a Guillain-Barre syndrome risk concentrated in the first five weeks, with attributable risk just under one per 100,000 vaccinations (Schonberger et al. 1979); comparable excess was not observed for the 1978-79 or 1979-80 influenza vaccines (Hurwitz et al. 1981; Kaplan et al. 1982). A Swedish registry cohort of 3.35 million Pandemrix recipients and 2.50 million unvaccinated people found narcolepsy HR 2.92 (1.78-4.79) among those aged twenty or younger, about four additional cases per 100,000 person-years, while the other specified neurologic and immune outcomes were largely near one (Persson et al. 2014). A post hoc case-cohort reanalysis of long-term dengue-vaccine trial follow-up detected increased hospitalized or severe dengue among baseline-seronegative recipients, a serostatus-specific harm rather than a uniform vaccine effect (Sridhar et al. 2018).

A same-estimator PCV7 positive-control attempt was also made with official CDC Active Bacterial Core surveillance. The serotype file contains 30,049 site-by-age-by-year rows for ten named sites over 1998-2024, while the national rate file contains 1,677 age-by-year rows over 1997-2024. Neither supplies birth cohort, year-specific site denominators, and a stable site-to-state catchment crosswalk simultaneously, so the state-by-cohort positive-control regression is fail-closed rather than estimated from incompatible dimensions. As a calibration only, national invasive-pneumococcal-disease rates fell from 1998-2000 to 2001-2003 by 72.2 percent below age one, 76.2 percent at age one, 43.7 percent at ages two to four, and 18.0 percent at ages five to seventeen. This age gradient is consistent with protection after PCV7 introduction but does not validate the exact autism contrast.

S38.5 Inferences allowed and forbidden

The observed patterns are inconsistent with a large generalized chronic effect that is proportional to coverage, appears in ordinary-risk populations, follows schedule additions, reverses after removals, compounds with additional vaccines, and recurs across the measured outcome systems. The theory audit separates the stronger autism contrasts from descriptive, falsification, and untestable modules; failure of several distinctive predictions does not convert every module into a causal estimate.

It remains forbidden to claim that the analyses prove zero risk for every individual, product, formulation, lot, simultaneous visit, days-long window, biomarker-defined syndrome, or ultra-rare subgroup. Mechanistic observations in autistic people do not by themselves identify the upstream exposure, and population-average nulls do not test every molecular pathway. The appropriate next studies require linked mother-child and sibling registries, exact product/dose/lot and visit histories, validated onset and clinical phenotypes, biomarker or genotype strata, and policy variation strong enough to change vaccination without changing diagnosis directly. These boundaries are part of the result rather than caveats appended after it.

S39. Combining information without confusing estimands

S39.1 Original observations and native policy contrasts

The primary synthesis models original observations with cohort-specific exposure mixtures (S54-S59 and S67). It does not average every available regression coefficient. U.S. PCV and RotaShield share a joint outcome likelihood; overlapping international systems are replaced or reported separately. Context and product parameters retain their own exposure definitions, outcome measurement assumptions, and bias allowances.

A native regression coefficient and an individual exposure-model ratio are not interchangeable. Multiplying a log coefficient by a scale factor changes the contrast but does not establish the counterfactual or correct a misaligned exposure. A population-level mixture contrast may be written as $Q(R;p_L,p_H) = [1+p_H(R-1)]/[1+p_L(R-1)]$, but an adjusted rollout score need not equal this raw contrast. Simulations must reproduce the actual adjustment and exposure paths before using its sign to infer the direction of an individual effect.

A common-direction average of noncommensurate policy coefficients is therefore not used for scientific inference. In particular, harmful exposure injections can generate a negative PCV rollout-spline coefficient (S52). Correcting dependence or choosing a different prior cannot fix that observation-model failure. The source-specific native results, conditional mixture fits, and their diagnostics remain available separately.

As a numerical commensurability check, the five-row re-expression uses ADDM on the IDEA 8.91-percentage-point contrast, yielding a synthetic ratio of 1.000 (0.956-1.060). This exceeds the six ADDM sites' observed 7.2-point uptake range and is not a causal common effect or an individual-risk bound. It has no weight in the original-observation synthesis. The shared-mean probabilities in S40 are calculated from that original-observation model, with all remaining identification assumptions retained.

S39.2 Individual-scale extrapolation calibration

A Bayesian individual-scale likelihood must predict the coefficient or outcomes produced by the actual design. For an exact two-population exposure contrast, one candidate observation model is

$$Q(R; p_L, p_H) = \frac{1 + p_H(R-1)}{1 + p_L(R-1)}.$$

The PCV7 rollout-spline coefficient is not this raw mixture contrast. Its archived fixed-share posterior, the individual-equivalent hierarchy, and later exposure redraws therefore do not supply a vaccine-risk posterior. The same nominal-share substitution requires validation separately for RotaShield and local MMR. These archived calculations document numerical algorithms, not current individual causal estimates. Direct source-based cohort-mixture and signed-score analyses are reported in S53.

Japan is not inherently unconvertible: its observed coverage trajectory can enter the full mixture model. Geographic mismatch and counterfactual trend assumptions limit causal interpretation; they do not forbid reporting conditional R estimates. Individual-study estimates retain their original RR, HR or OR scale and are not fused with policy contrasts merely because they are numerically near one.

S39.2a Individual-study search, pooling, and power audit

A dated public PubMed search through 30 August 2026 used eight specified query families and yielded 377 unique records. Of 189 records in the sensitive full-text queue, 132 were irrelevant title/abstract hits and five addressed methods or bias calibration. Forty-six records across forty-five data spines contribute 141 target neurologic rows and two allergic-reaction positive-control rows on their exact native scales. Two coefficients are genuinely interval- or access-blocked: Kawai et al. (2014; PMID 24497128), whose abstract reports no increased seizure, Guillain-Barré syndrome, encephalitis, or anaphylaxis risk after 2012-2013 influenza vaccination but omits the complete adjusted effect table, and Tartof et al. (2014; PMID 25285882), whose abstract reports comparable outcomes after vaccine-associated and non-vaccine-associated febrile seizures but omits the adjusted comparative interval. Publisher access prevented primary-table verification, so neither coefficient is back-filled from a secondary review or treated as a null. The exact-effect ledger contains 168 native-scale coefficients from sixty-five PubMed records and sixty-one specified spines: 166 calibration rows and two separately partitioned allergic-reaction positive controls. No independent ASD estimate was identified. Forty spines contain multiple rows, generating 302 explicit within-spine row-pair relations; the ledger also records ten cross-label spine warnings. Every row has zero policy-synthesis weight and automatic cross-scale pooling disabled. Repeated outcomes, components, estimators, VSD/Sentinel/Korean/register spines, and repeated follow-up are encoded rather than counted as replications. WHO's 31-study 2010-2025 review and the 2026 BMJ systematic review of fifty-nine aluminum-adjuvant studies were separately cross-walked. This is a structured dated public search and extraction audit, not a duplicate-screened PRISMA review of proprietary databases.

WHO's evidence review assessed thirty-one primary studies and five meta-analyses published from January 2010 through August 2025 (WHO 2026), and WHO's Global Advisory Committee on Vaccine Safety reaffirmed its 2002, 2004 and 2012 conclusions (GACVS 2025). The 2026 systematic review of fifty-nine human aluminum-adjuvant studies, with searches updated to 27 November 2025, located only three studies reporting autism: two ecological correlations judged at critical risk of bias and the Danish aluminum cohort. None of its eleven randomized trials, mostly reactogenicity comparisons in adults, reported an autism or other neurodevelopmental outcome, and the review identified uncommon, localized, usually self-limited nodules or granulomas after some aluminum-adjuvanted vaccines

(Doyon-Plourde et al. 2026). Its autism conclusion therefore rests on the Danish cohort. Reverse selection is documented directly: in the Vaccine Safety Datalink, children diagnosed with autism were less likely than other children to be fully vaccinated for vaccines due at ages four to six years (adjusted RR 0.87, 0.85-0.88), and their younger siblings were also less often fully vaccinated (Zerbo et al. 2018), which biases naive vaccinated-versus-unvaccinated contrasts toward an apparently protective association. The Danish Calmette trial's developmental comparison, an adjusted difference of -0.7 points (-3.7 to 2.4) in 12-month Ages and Stages scores after BCG, is a complete-case analysis with 81 percent response (Kjærgaard et al. 2016).

No mixed RR/OR/HR omnibus pool is reported. This section describes a restricted REML calibration; S67 reports the native-scale Bayesian families with explicit bias allowances. The one-spine-per-study MMR cohort sensitivity combines three person-disjoint cohorts: Madsen (RR), Hviid (HR), and Burstain (HR), with the latter's reported 99-percent interval converted to the common 95-percent scale. Madsen and Hviid use nonoverlapping people but share Danish national-register infrastructure, so they are not described as fully independent. Treating RR and HR as approximately commensurate gives 0.961 (modified-Hartung-Knapp 0.877-1.052; prediction interval 0.877-1.052). The Smeeth and Uno case-control ORs do not enter this cohort calibration; their separate OR family is reported in S67. Four geographically distinct prenatal-influenza cohort HRs—Zerbo, Ludvigsson, Becerra-Culqui, and Neeman—combine to 1.019 (0.919-1.130; prediction interval 0.860-1.208). Fixed-effect, DerSimonian-Laird, Paule-Mandel, and REML estimates differ by less than 0.003 in either family. Positive assumed sampling correlation up to 0.99 between the two Danish cohorts does not create a harmful point estimate. This paper does not add a new study-level risk-of-bias score; the cohort layer is explicitly observational, receives zero causal weight, and is read alongside the appraisals in the cited WHO, Cochrane, and AHRQ reviews.

For this restricted observational calibration, with small-k modified-Hartung-Knapp inference, the prospective 80-percent MDE is RR 1.127 for the MMR cohort family and 1.149 for prenatal influenza. The observed MMR interval nevertheless excludes RR 1.10 because its point estimate is below one; the prenatal-influenza interval excludes RR 1.20 but not RR 1.10. Normal-likelihood Bayes factors comparing the point null with positive half-Normal alternatives favor the null by 6.8, 13.0, and 24.7 to one for MMR prior scales RR 1.05, 1.10, and 1.20; corresponding prenatal-influenza values are 1.17, 1.89, and 3.39. A small-k Student-t likelihood reduces the MMR values to 3.0, 5.0, and 8.8 and the prenatal values to 1.14, 1.78, and 3.13. The heavier-tailed sensitivity is the more cautious reading. Both sets are prior- and likelihood-conditional observational calibrations, not causal Bayes factors.

The Danish shared-system covariance sensitivity is conditional on an assumed correlation, not estimated participant overlap. Nonnegative common-effect weights are computed by enumerating the feasible active sets and independently checked against optimality conditions and a rescaled numerical optimizer. Across correlations from zero to 0.99, the MMR association remains 0.961 after rounding, with normal-theory interval endpoints spanning 0.922-1.003. Every interval includes one. At correlations of 0.50 or higher the less precise Madsen row receives zero weight. This is a common-effect numerical sensitivity, not the modified-Hartung-Knapp primary interval, and it does not remove shared structural bias. Unconstrained GLS can produce negative weights and artificially narrow intervals at high assumed correlation; those rows are not interpreted as evidence of protection (file 2810).

S39.3 Schedule-interaction hierarchy

Interaction coefficients estimate departure from log additivity and are not mixed with main effects. The core interaction set retains the directly measured U.S. PCV-by-baseline schedule estimate, its ADDM analogue on the same (+10-point) by (+10-point) contrast (1.018, 0.280-3.700), a dependence-adjusted summary of six U.S. vaccine-pair estimates, a corresponding six-pair England upper-tier local-authority summary, and the Wales rotavirus-by-PCV estimate. Within-panel pair summaries use positive inverse-variance weights with rho = 0.50 design-effect inflation. Scotland, Northern Ireland, Sweden, and England nested national additions are excluded because pretrend, placebo-cutoff, age-period-cohort, or support diagnostics fail. Oregon enters only an expanded sensitivity because its additive autism-rate interaction must be approximated as log RR by dividing by the weighted outcome mean.

The normal-normal reference hierarchy gives a posterior-median interaction RR of 1.009 (0.953-1.074), with 30.2 percent probability above 1.02, 7.35 percent above 1.05, and 0.88 percent above 1.10. The predictive interval is 0.900-1.137. Sampling independence and maximal dependence give mean intervals of 0.954-1.072 and 0.951-1.077. Doubling design-bias allowances gives 1.008 (0.941-1.085); adding the approximate Oregon family gives 1.002 (0.943-1.057). These describe the specified ecological interaction contrasts under the model, not all same-child biological interactions or total schedule effects.

The hierarchy is a descriptive stress test across differently scaled ecological joint contrasts; the directly measured PCV-by-baseline-schedule coefficient remains the interpretable anchor. Its archived point estimate is RR 1.001 (0.990-1.011); 499 Kish-effective joint NIS redraws give a conservative interval of 0.989-1.012 and a Rubin interval of 0.989-1.013. The interval width rises by only 9.4 percent. Classical and fixed-effect-partial interaction reliabilities are 0.914 and 0.856, the observed target VIF is 2.17, and deleting one linked cohort at a time gives RRs 0.998-1.004 except for deletion of the lone pre-rollout cohort, whose interval is 0.957-1.047. Under a state-AR(1) simulation with latent rather than published exposure, null rejection is 5.6 percent, power is 38.6 percent at RR 1.02 and 93.2 percent at RR 1.05, and the empirical 80-percent MDE is 1.043. These power results apply to a homogeneous ecological departure from log additivity on this measured contrast, not exact same-visit or rare-subgroup synergy.

The model does not identify a biological interaction for simultaneous administration, exact visit spacing, a rare susceptible subgroup, or a combined product formulation such as MMRV. Nor can the national schedule itself identify a state-by-cohort antigen-count effect: national dose additions are absorbed by cohort fixed effects, while state schedule-completion measures vary uptake rather than the assigned antigen count. The MMRV febrile-seizure positive control remains an example of a formulation-specific acute interaction detectable in the appropriate days-long risk window. Figure S41a makes the design logic explicit: a harmful incremental addition should create an upward cohort step, a removal should reverse that prediction, and an interaction requires off-diagonal schedule support beyond the A-to-A+B comparison. Continuing identification growth after schedule saturation is inconsistent with measured schedule drift as the principal engine.

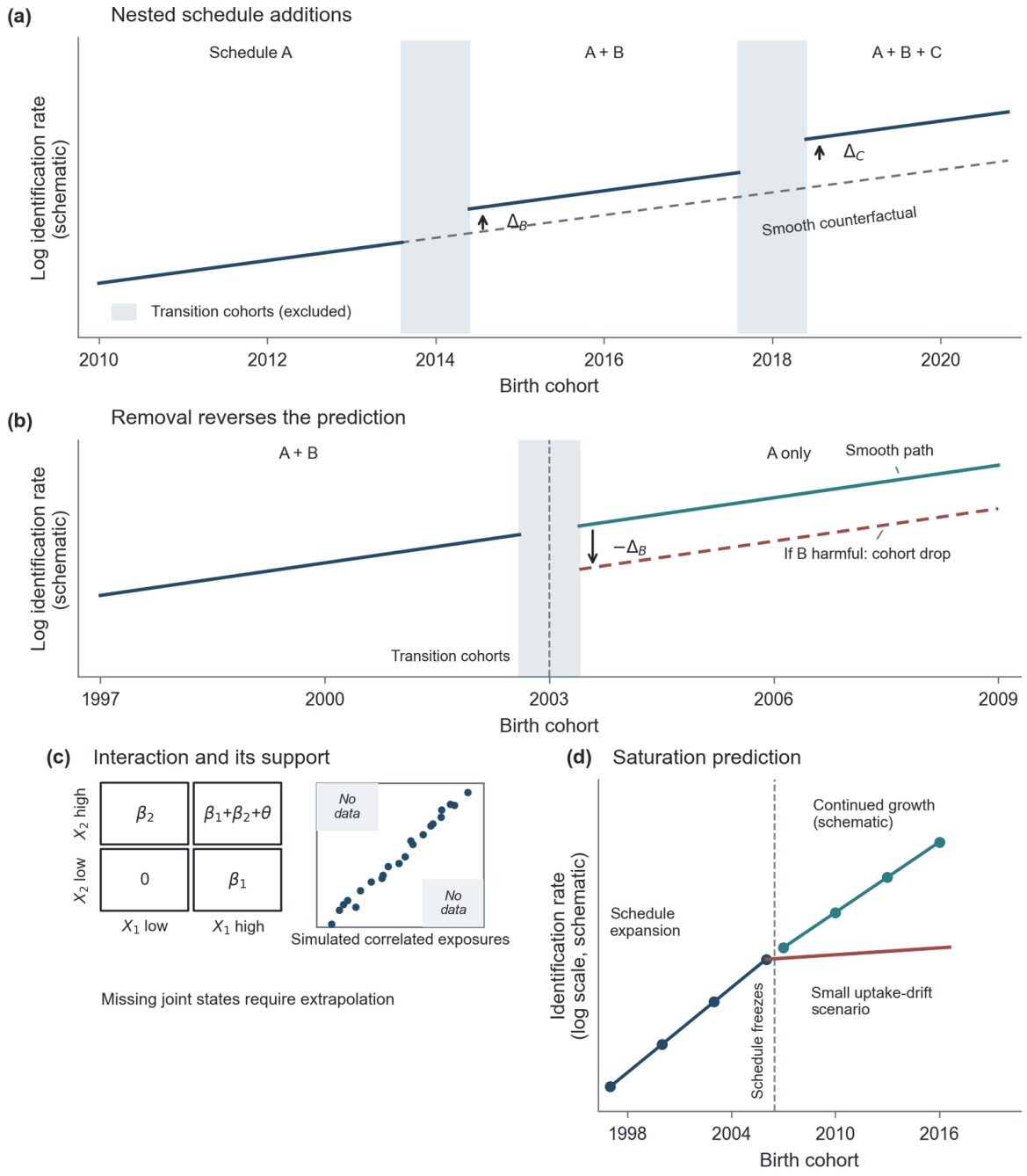


Figure S41a. Design logic for additions, removals, schedule interactions, and saturation. All paths and exposure points are schematic, not additional observations or effect estimates. Under a stable counterfactual, candidate harmful effects imply sign-reversing cohort steps at additions and removals. Missing schedule combinations require extrapolation; the illustration does not classify the support of any particular empirical panel.

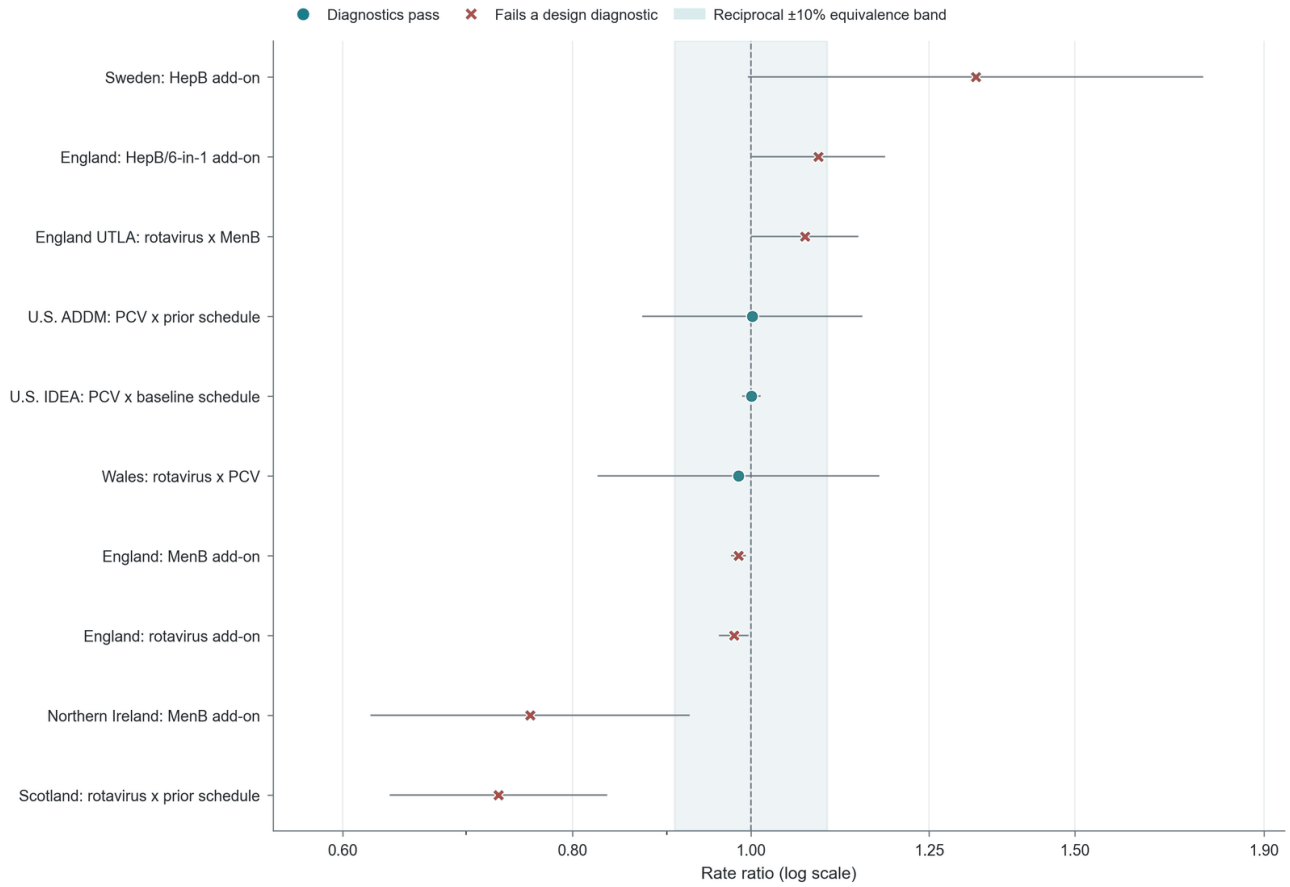


Figure S41b. Addition-on-top-of-schedule estimates and interaction diagnostics. The gate column distinguishes supported joint contrasts from estimates that fail timing, support, or age-period-cohort diagnostics; a failed gate is not converted into null evidence.

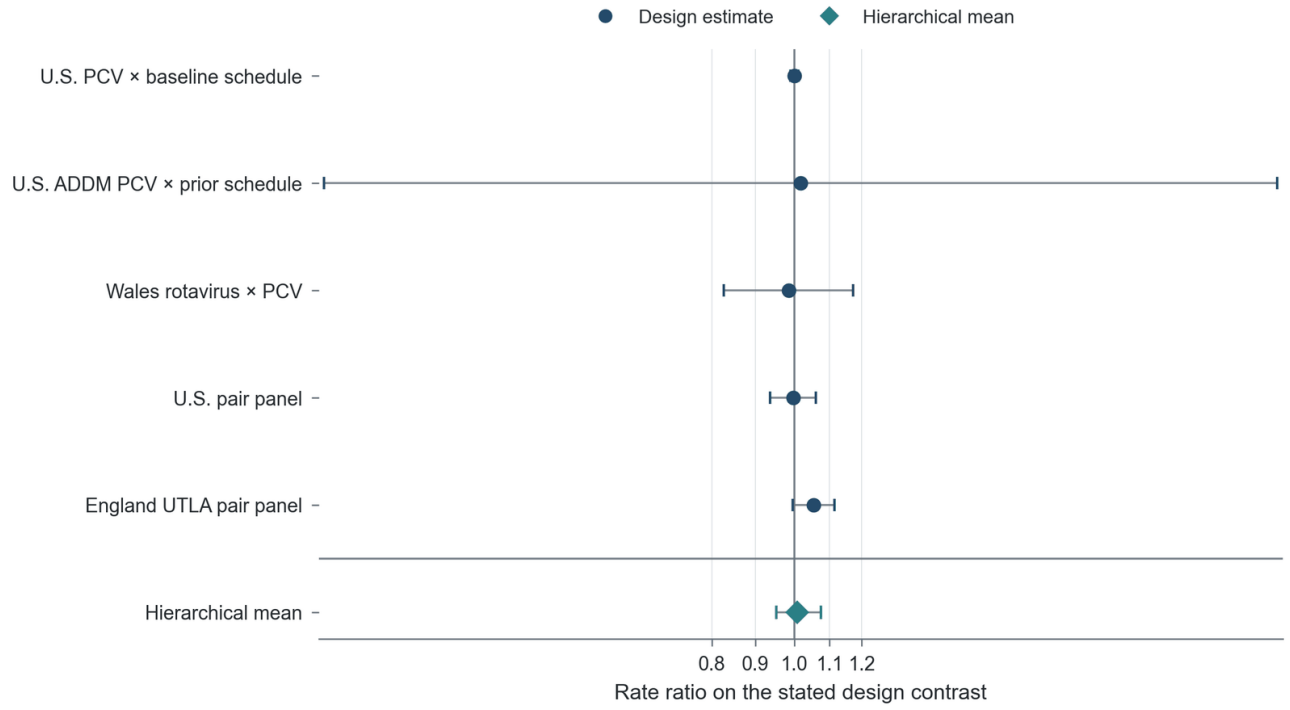


Figure S41c. Schedule-interaction hierarchy. Failed timing, age-period-cohort, and unsupported nested-schedule designs receive zero weight; the estimand is departure from log additivity on each supported joint contrast.

S39.4 Reproducibility and interpretation rules

The current original-observation model is implemented in `processed/global_hierarchy.py`; its authoritative reference and sensitivity summaries are `global_integration/hierarchy/hierarchy_final_p18.json`. Files 3320 and 3309 provide the source-linked synthesis and checked-design tables. The shared-mean probability suite is generated by `processed/current_mean_probabilities.py`. Historical prediction-stress and coefficient-translation outputs remain archival artifacts, not current inferential inputs; their obsolete probability-display entry points require an explicit archival opt-in.

Native contrasts, context parameters, product means, interactions and the overall shared mean answer different questions. Shared observations enter a joint likelihood or a single selected context; they are not multiplied by the number of specifications. Failed exposure mappings cannot be repaired by assigning a hierarchical prior. Lower and upper tails are both reported, but a negative posterior component is not automatically evidence of a protective causal effect.

S40. Negative, practically null and positive shared-mean probabilities

S40.1 Target and null definitions

This section concerns the shared mean mu of the fifteen-context, fourteen-block original-observation hierarchy in S59. It does not concern the probability that every vaccine effect is zero. Product and context heterogeneity remain active even when mu is exactly zero, so positive and negative context effects can coexist under the shared-mean null.

For a continuous prior, the probability of exactly mu = 0 is zero. A practical-null classification instead uses a reciprocal margin c: negative means exp(mu) < 1/c, practically null means 1/c <= exp(mu) <= c, and positive means exp(mu) > c. We report c = 1.01, 1.02, 1.05 and 1.10. At the five-percent margin the lower threshold is 0.95238, not 0.95.

The saved likelihood is evaluated over a = mu + b, where shared bias b has a Normal prior with mean zero and standard deviation 0.15. The continuous reference prior for mu is Normal with mean zero and standard deviation log(2). Conditional on a, mu is integrated analytically. The six product/context heterogeneity components retain their stated prior weights.

S40.2 Continuous posterior and prior sensitivity

Under the reference prior, exp(mu) is 0.957 (95% credible interval 0.599-1.517). The probability of a negative shared mean is 57.5 percent and of a positive mean is 42.5 percent. At the reciprocal five-percent margin, the negative, practical-null and positive probabilities are 49.1, 16.3 and 34.5 percent. At the ten-percent margin they are 41.2, 31.3 and 27.5 percent. These are broad distributions, not tight individual-risk bounds.

Holding shared-bias SD at 0.15 while halving the mu-prior SD gives 0.968 (0.650-1.437), with 26.2 percent probability above 1.10 and 19.0 percent in the reciprocal five-percent region. Doubling the mu-prior SD gives a grid-based diagnostic of 0.953 (0.581-1.545); its numerical reliability has not been independently established under the new core weighting, so it is not used as a validated tail bound. These fixed-bias changes differ from S59's jointly scaled mu-and-bias prior scenarios.

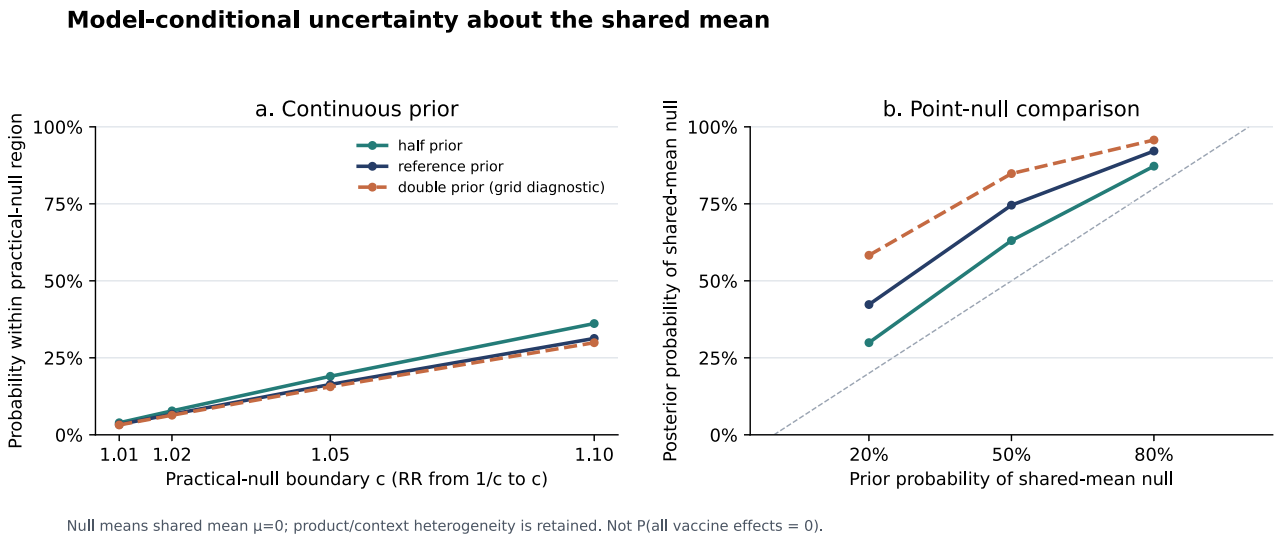


Figure S68. Shared-mean probability sensitivity. The negative, practical-null and positive regions use reciprocal ratio margins, with shared-bias SD fixed at 0.15. Point-null comparisons fix only the shared mean at zero; product/context heterogeneity remains. Posterior null probabilities depend on prior null mass. The doubled-prior curve is a grid-based numerical diagnostic, not a validated tail bound. These are not probabilities that every vaccine effect is zero.

S40.3 Point-null model comparison

The point-null model fixes mu = 0 but retains shared bias and product/context heterogeneity. Against the reference continuous prior, the Bayes factor favoring this shared-mean null is 2.932. With prior null masses of 0.20, 0.50 and 0.80, posterior null probabilities are 42.3, 74.6 and 92.1 percent. Thus a high posterior null probability can be largely prior-driven. Halving the continuous-prior SD reduces the Bayes factor to 1.708 and gives posterior probabilities of 29.9, 63.1 and 87.2 percent at the same prior masses.

The alternative is a continuous distribution of positive and negative shared means, not an all-harm model. This comparison cannot establish that all vaccines are harmless, that all individual effects are zero, or that any particular outcome lacks a causal effect. Interaction effects and cumulative exposure have different estimands and do not inherit these probabilities.

S40.4 Numerical checks and reproducibility

The suite reproduces the reference hierarchy's median and interval exactly from the saved likelihood grid. A half-grid check (121 rather than 241 points) changes the directional probability by less than 3 x 10^-9, the five-percent practical-null probability by less than 6 x 10^-9, and the log Bayes factor by less than 4 x 10^-8. These checks address quadrature on the saved grid, not exposure error or causal identification. The reference minimum relevant core effective sample size is 106; separate low-effective-size sensitivity settings remain flagged in S59.

The files `mean_probabilities.json`, `mean_probabilities.csv`, `mean_pointnull_probabilities.csv` and `mean_probabilities_grid_sensitivity.csv` in `project_results/global_integration/` contain the full prior, margin and null-mass grids, source hashes and numerical diagnostics. Focused tests check exact reference agreement, probability partitioning, the point-null Bayes identity and quadrature stability.

S41. Global schedule-event accounting

The date-locked schedule-event analysis frame contains 3,379 records: 3,358 WHO-derived records and twenty-one independently verified Chilean and Latvian events that are not duplicates of another episode. Ten source records clarify or enrich an existing episode without creating another event. File 2163 documents the thirty-one reconciliation decisions; files 2164-2167 rebuild and validate the analysis frame. The unit is an event record. Multiple records can describe a shared cutoff, component, jurisdictional mandate family, or national recommendation, so neither the row count nor a large set of regressions is an independence count. Separate curated and subnational ledgers include source-specific events such as the 2014 Saguenay-Lac-Saint-Jean MenB campaign; those nonadditive ledgers are not included in the 3,379-row denominator.

The primary-source catalog records forty-nine antigen- or phase-level sources mapping to eighteen linked policy episodes. Six are source-detail supersessions or expansions of an existing canonical episode, leaving forty-three active antigen/event records and twelve net linked episodes beyond the date-locked analysis frame. The public inventory contains 3,422 active canonical records and 3,391 linked episodes under the explicit legacy-singleton accounting rule; the analysis frame contains 3,379 records. Seventy-two apparent WHO/JRF withdrawal or reinstatement transitions are classified as `probable_unverified_jrf_transition`; their strict-date and registry-ready flags are false pending primary national confirmation. None is causal-ready, and missing outcomes or failed gates are not coded as null effects (files 2168, 2170, and 2171).

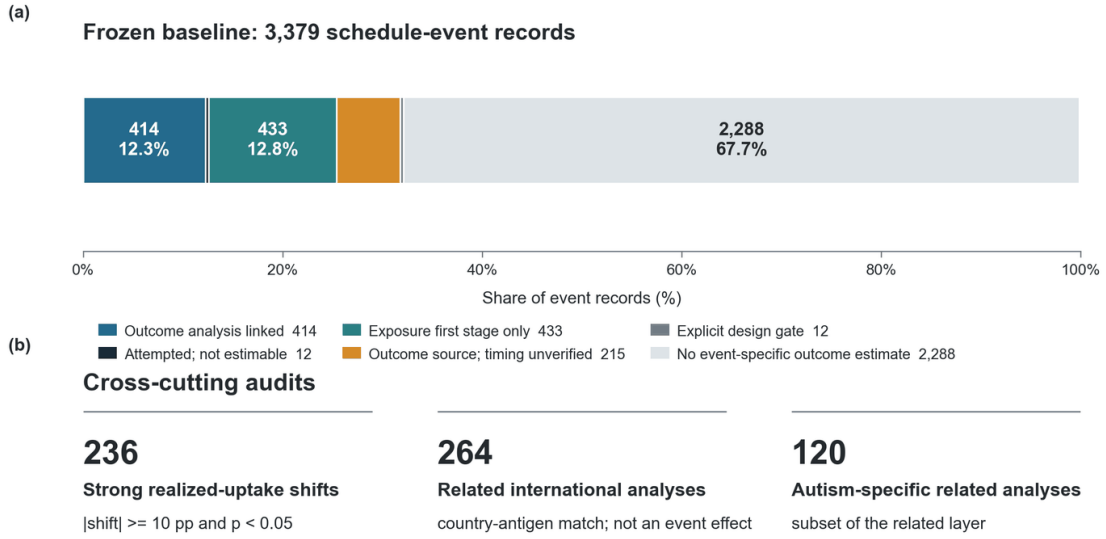


Figure S43. Complete schedule-event accounting. All 3,379 registry records remain visible and are partitioned by event-specific testing status; first-stage and international-result matches are cross-cutting audits rather than independent effect estimates.

S41.1 Exposure tests for the complete registry

Every original registry row is passed through the same event-window exposure procedure. Product-compatible WHO coverage endpoints are available for 2,600 unique country-indicator-year-direction cells. Of these, 475 have sufficient observations on both sides of the listed date for a HAC interrupted-series first-stage estimate, and 235 show an absolute level shift of at least ten percentage points with p below 0.05. The twenty-one new verified events add one more strong first stage, for 236 in the augmented universe. Among the 475 observed two-sided tests, 260 reject the pre-event trend diagnostic at p below 0.05. That failure does not erase the event record; it prevents the listed date from being treated as an unqualified discontinuity. Files 2135-2141 and 2180 contain the coverage panel, unique-window estimates, row-level ledger, validation checks, and audit report.

S41.2 Event-specific outcome-testing states

Files 2150-2156 and 2179-2184 assign every event record one mutually exclusive outcome-testing state. Outcome analyses are linked to 414 records, including diagnostic and zero-weight analyses; twelve further records were actively attempted but cannot be estimated with the acquired outcome support. Another 433 have an observed event-window exposure first stage but no event-specific outcome estimate. For 215 records a country-level outcome source exists, but the historical implementation date or affected age has not been verified well enough to define treated cohorts. Twelve new Chilean records hit explicit outcome-support or counterfactual gates. Five further records have a candidate outcome source but an unresolved reanalysis gate, and the remaining 2,288 records have no compatible event-specific outcome estimate in the acquired public data.

The final category is a data-compatibility result, not a set of omitted regressions. For 1,867 of the 2,288 records, no acquired repeated outcome dataset maps to the country; for 421, a mapped outcome source exists but its calendar span does not cover the necessary lead and post-exposure cohorts. All 2,288 also fail the event-specific exposure gate: 1,992 lack sufficient WHO observations on both sides of the listed date, 286 have no compatible WHO product or dose indicator, and ten subnational events require unavailable subnational exposure series. These dimensions overlap rather than sum. Historical implementation date or affected age is not strictly primary-source verified for 2,278 of the records. No event has verified timing, overlapping acquired outcomes, a usable exposure contrast, and a credible counterfactual but was simply left unanalyzed.

“Linked” is not synonymous with “causal.” Three hundred forty-one U.S. state mandate rows collapse to thirteen mandate families. All thirteen have model-estimated lead diagnostics, seven have adequate power for RR 1.10, three are post-outcome timing placebos, no family passes the frozen causal gate because mandate timing is not a measured uptake first stage, and the only family with a Holm-adjusted p value below 0.05, a single-state late-childhood HepB family (Holm $p = 0.0088$), fails its lead test at $p = 1.8e\text{-}76$. Twenty-five national U.S. non-mandate events receive linear and quadratic interrupted-series screens, but all retain zero causal weight because there is no untreated national counterfactual and the estimates are unstable to curvature or arbitrary cutoffs. Canada's staggered rotavirus introductions link eight provincial records to an independent PHAC CCDSS outcome panel. The naive event study is positive but fails its lead tests overwhelmingly; the province-trend cohort-saturation specification is null. Chilean and Latvian national interrupted-series tests are likewise retained as low-tier, zero-weight diagnostics. These rows close the testing denominator without manufacturing independent evidence.

S41.3 Global published-prevalence event screen

The CDC census of published population-based autism prevalence studies supplies one additional, deliberately weak way to interrogate national events. Script 2192 attempts every strictly verified national event in the augmented universe. A linear and quadratic cohort-trend screen is estimated only when the same country has at least four study records before and four after the first affected cohort within a fifteen-cohort window. Method and diagnostic-criterion families enter as controls, standard errors use HC3, and all fitted specifications share one Holm family.

All 110 currently classified strictly verified national event records are attempted: 67 belong to the frozen 3,379-record frame and 43 enter through the current closeout layer. Twenty-nine have sufficient two-sided study support, producing fifty-eight fitted specifications; eighty-one fail the support gate. None survives Holm correction. The same published-prevalence study records are reused across nearby cutoffs, some estimates are extremely unstable to linear versus quadratic curvature, and the observations combine different places, ages, ascertainment systems, and diagnostic criteria. Every row therefore has zero causal weight. Files 2192-2196 contain the attempts, estimates, duplicated event-study panel, validation checks, and audit. The value of this exercise is denominator closure and model-failure visibility, not another pooled null.

S41.4 Related international analyses are crosswalked, not relabeled

Some countries have extensive outcome analyses that are informative about coverage gradients or administrative associations but do not identify a particular introduction or withdrawal. Script 2185 reviews 11,144 distinct international statistical-result rows from sixty-six files in the current harmonized archive. Because a source row can concern more than one antigen, the script emits 12,158 row-to-family memberships in 1,099 fine-grained country-source-exposure-outcome families across thirty-two countries; no membership or family is uncategorized. Of 11,657 ratio-scale memberships, 11,620 have valid intervals and 37 are excluded under the specified interval contract. No family is source-marked causal-ready. The country-antigen crosswalk associates this related layer with 264 frozen event records, 120 of which have at least one autism-specific related family. Files 2185-2190 preserve the source, exposure class, outcome scope, membership count, interval status, and matching event records. These memberships and within-source specifications are not independent studies, and every link carries the explicit role `related_country_antigen_analysis_does_not_by_itself_test_this_event`.

S41.5 Australia public-data extension

Sources and outcome definition

The AIR acquisition manifest enumerates 428 distinct official PHN/SA3 spreadsheet links across nineteen publication pages. At the recorded retrieval, 129 were accessible and 299 returned HTTP 403. File 2144 preserves the URLs, retrieval status and checksums; this is a documented acquisition inventory, not a claim that inaccessible files were acquired. The June 2026 NDIS release contains twenty-eight diagnosis reporting dates and twenty-seven SA3 participant-count dates. Diagnosis is available longitudinally by state, but not cross-classified by SA3. SA3 total enrollment cannot supply a local autism numerator.

The NDIA diagnosis data rules distinguish one primary disability per participant from nonexclusive reported diagnoses. Primary disability reflects greatest functional impact; the agency does not independently validate diagnoses. State codes begin in June 2020. We sum primary counts, not reported counts. The administrative Autism group includes F84.0, F84.2 (Rett syndrome), F84.5 and unspecified records. The count audit finds 22,106 state-diagnosis-date rows, 8,449 suppressed primary counts, and no duplicate full diagnosis keys. The parser rejects unrecognized count tokens.

Model and diagnosis sensitivity

The state composition model relates the logit of the primary-autism share to age-two AIR schedule completion, with state and year fixed effects. The reference fit weights by total primary-disability counts and uses a state-clustered covariance matrix with a t(7) reference distribution. Ratios are approximate enrollment-share ratios per ten percentage points of schedule completion, not individual autism risks. The contemporaneous estimate is 1.092 (95% CI 0.893-1.306); future exposure gives 1.023 (0.767-1.312), lag one 1.082 (0.937-1.234), and lag two 1.031 (0.855-1.217). File 2148 retains suppression replacements of zero, five and ten, assessment ages one, two and five, and both weighting schemes.

Holding the denominator fixed and restricting the numerator to F84.0/F84.5 gives a contemporaneous ratio of 1.091 (0.890-1.307); F84.0 alone gives 1.108 (0.900-1.333). The complete definition grid crosses three definitions, three suppression replacements, four exposure timings and two weighting schemes (72 overlapping fits). Nine intervals are nominally positive: all unweighted lag-one fits across the three definitions and replacements, with ratios 1.134-1.162 and p-values 0.016-0.030. None is nominally negative. A post hoc Holm adjustment across all 72 fits yields no rejection (minimum adjusted $p = 1.000$). This does not establish the validity of the small-cluster p-values or negate the consistent unweighted pattern.

State deletion does not explain away that pattern. With replacement five, all eight deletions retain nominally positive unweighted lag-one intervals for all three definitions. Deletion ratios range from 1.105 to 1.173 for the administrative group, 1.105-1.169 for F84.0/F84.5, and 1.113-1.203 for F84.0 alone. Weighted administrative-group lag-one ratios range from 1.037 to 1.136, with no interval excluding one. The full influence grid contains 12 full-sample and 96 deletion fits, covering both timings, weightings and all definitions. Deleted-state fits use t(6). Deletion changes the population being summarized, and weighting differences may reflect heterogeneity as well as influence.

Denominator checks and limits

The privacy rules allow modification of low cells and related data while leaving published aggregate totals unchanged. Replacements of zero, five and ten are sensitivity scenarios, not guaranteed bounds on true counts. Across

200 matched state-quarters, midpoint-substituted diagnosis totals divided by separately published SA3 totals range from 0.99995 to 1.09777. Through December 2024, the range is 0.99995-1.04192, with the largest difference in the Northern Territory in September 2022. SA3 counts concern active participants at their latest residential address and include an Other category for missing and small-area records.

Allowing every visible suppressed cell to vary independently from zero to ten can reconcile all matched totals through 2024, but not five Victorian quarters in 2025-2026. In June 2026, midpoint totals are 210,873 versus 192,093; even the minimum diagnosis sum exceeds the maximum SA3 sum by 18,655. The gaps coincide with published drops in Wyndham (9,742 in March 2025 to 1,016 in June) and Geelong (9,821 in December 2025 to 1,004 in March 2026). Both rows remain present. A fresh official download matches all 9,385 archived rows byte-for-byte, ruling out local CSV transformation as the source of these values. The cause remains unresolved. We do not infer replacement counts or substitute the SA3 denominator. These later quarters can enter lagged outcomes but are outside the contemporaneous 2020-2024 window.

The analysis retains the robust unweighted association without interpreting it as a policy-induced vaccine effect. Endogenous schedule completion, enrollment selection, cohort mismatch, scheme access changes, the unresolved denominator discrepancy and unvalidated small-cluster inference prevent causal attribution. Neither the sensitivity fits nor the alternative NDIS tabulation add independent participant evidence.

Reproduction

A fixed-design simulation checks the conventional CR1/t(7) test under independent-state Gaussian logit errors. It crosses contemporaneous and lag-one designs, both weightings, within-state AR(1) correlations of zero, 0.5 and 0.9, and equal versus residual-scaled state variances. With 10,000 draws per scenario, null rejection ranges from 2.04% to 10.13%, rather than uniformly 5%. Conventional power for an enrollment-share ratio of 1.20 per ten percentage points ranges from 49.11% to 99.58%; it is not size-adjusted. Calibrating each scenario's critical value on 5,000 null draws and evaluating the other 5,000 gives null rejection of 4.36% to 6.26% and power of 57.6% to 98.2%. These ranges depend on assumed error processes and fitted noise scales, not only sample size. The scenario-specific calibration is not a corrected p-value for the observed data. Endogeneity, cross-state dependence, suppression uncertainty and identification bias are outside this simulation; the target remains enrollment composition, not individual autism risk.

The age-design module contains the raw-cell audit, 72-fit definition grid, state-deletion grid, denominator reconciliation and SA3 discontinuity audit. A separate CSV and matrix implementation reconstructs aggregation, annual matching, weighted least squares and the CR1 state-cluster covariance for all 72 definition fits; 504 reported numbers agree within 3.4e-11. This checks implementation, not causal assumptions. Run `run_extension_reproduction.py --run --groups australia` from the module path specified in the reproduction instructions. `reacquire_australia_sa3.py` separately verifies the official download; it is not part of the offline-oriented analysis runner.

S41.6 District-age enrollment comparisons

Eight official NDIS participant-budget snapshots cover September 2024-June 2026. They contain 5,707 local age-specific autism rows with matching all-disability rows, not 5,707 independent populations. Support class ALL avoids adding overlapping support categories. Childhood bands are combined into ages 0-14 across the December 2024 age-band change; adults are grouped at ages 25-64. Every component must be present. The resulting panel has 1,256 district-date-age rows, with 168 missing-component groups recorded separately rather than filled with zero. Seventy-two named district labels support both ages at all eight dates and match the official 2021 LGA-to-district mapping. Matching names does not establish unchanged boundaries or exposure histories.

On this common support, midpoint-substituted autism shares of active-plan enrollment rise from 49.96% to 60.97% among children and from 12.76% to 17.22% among adults. The descriptive child-minus-adult change is 6.56 percentage points. Allowing counts to vary within their literal published suppression tokens gives endpoints of 4.42-8.69 percentage points. These endpoints are not confidence intervals or guaranteed bounds under the agency's wider privacy modifications. Unknown tokens are rejected; thresholds such as less than 1,001 are not replaced by five. An independent raw-CSV aggregation exactly reconstructs all sixteen pooled age/date margins.

A relevant administrative process is documented in the NDIA's Annual Financial Sustainability Report 2024-25, Section 5.3.3, printed page 96: increased eligibility-reassessment activity accompanied increased transitions from developmental delay to autism or intellectual-disability classifications in 2024-25. This does not quantify how much of

our observed change it explains. It does show why adult differencing need not remove child-specific changes in recorded primary disability. These snapshots do not identify a vaccine-exposure contrast, and adults are not assumed to be historically unexposed. The positive descriptive contrast remains visible; it is neither a vaccine effect nor a null vaccine result.

The corresponding numerator sensitivity holds the childhood denominator and observation dates fixed. In published national totals, autism alone rises from 49.70% to 60.81% of age-0-14 enrollment, a change of 11.11 percentage points. Combining the disjoint primary categories autism, intellectual disability, developmental delay and global developmental delay yields 90.19% to 90.89%, a change of 0.71 points. These national cells are numeric rather than suppressed. The attenuation is compatible with reclassification within the broader group but cannot identify individual transitions or explain the entire change. Adding intellectual disability alone leaves a 10.72-point increase, so the developmental-delay categories drive the attenuation. These are overlapping analyses of the same administrative system, not independent replications.

At district level, requiring every component category at every date reduces common support to 45 districts. Autism alone rises 10.96 points there. The broad numerator's midpoint change is -8.98 points, but literal suppression-token endpoints span -20.29 to +2.32 points. That sensitivity prevents a robust directional conclusion for the broad district result. Missing categories are not treated as zero, and the national totals are not substituted for the district cells. None of these descriptive changes is a vaccine-effect estimate.

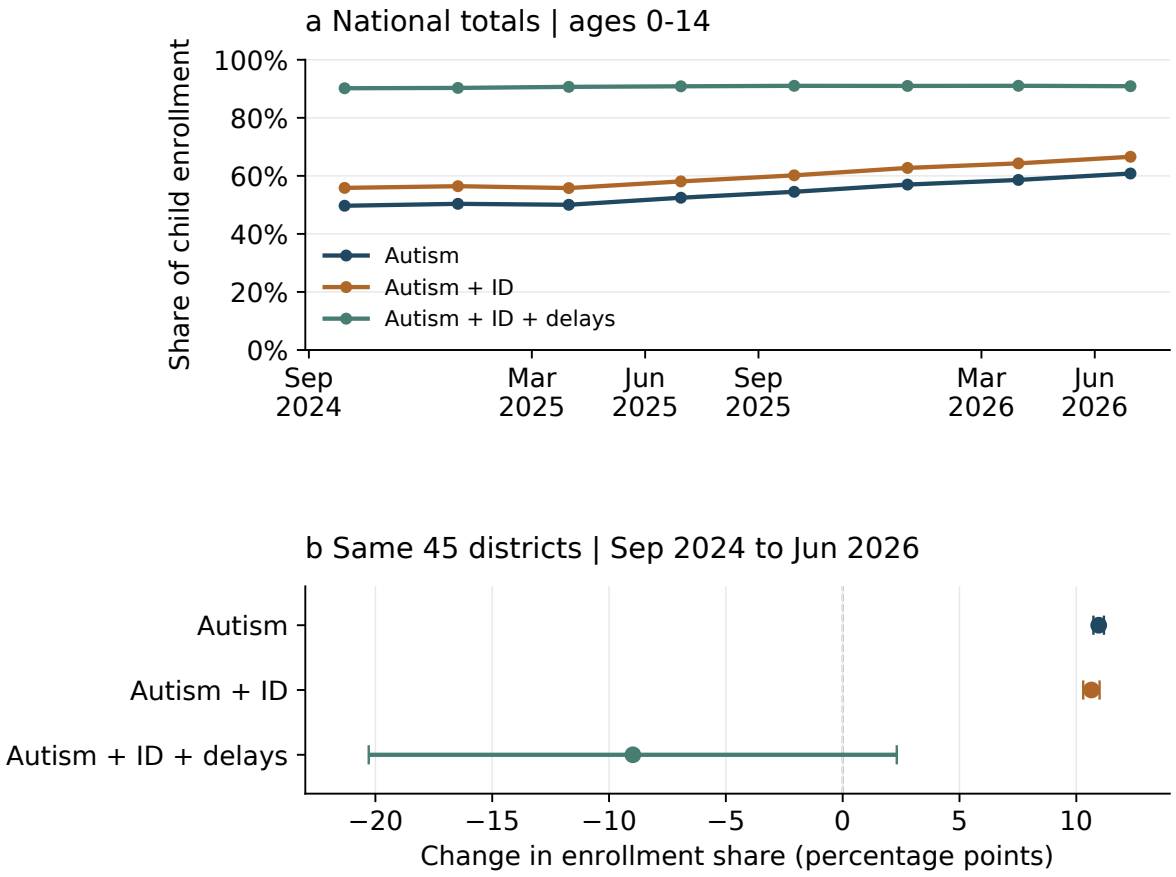


Figure S54. Recorded disability composition and numerator sensitivity. Panel (a) includes all eight national NDIS snapshots. ID denotes intellectual disability; delays includes developmental delay and global developmental delay. Panel (b) holds geographic support fixed across all categories and dates. Points use token midpoints; bars show literal suppression-token sensitivity, not confidence intervals or guaranteed privacy bounds. National and district summaries overlap and are not independent replications. Outcomes are shares of active-plan enrollment, not population incidence or vaccine effects.

S41.7 Cohort-specific coverage and joint-receipt bounds

The NCIRS Annual Immunisation Coverage Report 2024, Tables A3-A10, provides 1,360 coverage cells: 560 childhood cells across 31 Primary Health Networks and overlapping state/national totals, and 800 adolescent cells by state, reported sex, birth cohort and vaccine category. Indigenous-population tables overlap the all-participant tables. These are AIR exposure measurements, not independent health-outcome observations. All extracted rows agree across PDF table-geometry and text-order checks, and a fresh-download reproduction matches their normalized hashes.

The childhood tables retain birth years 2023, 2022 and 2019 and seven assessment ages. For matched birth cohorts and coverage algorithms, median unweighted PHN coverage increases are 3.9 percentage points from 9 to 12 months, 4.0 from 21 to 24 months, and 7.5 from 51 to 60 months among all children. These differences describe receipt timing; they are not health effects or comparisons of one constant schedule across all seven ages. The report's printed regional values are retained, including unusual North Coast and Northern Sydney 60-month values that require cross-release reconciliation before substantive interpretation.

The adolescent tables measure HPV initiation and receipt of adolescent DTP-containing and MenACWY doses through 31 December 2024, using AIR data as at 2 February 2025. Across 160 common state-by-sex-by-population-by-birth-cohort cells aged 15-19, the proportion receiving all three categories is bounded below by the larger of zero and the sum of the three marginal proportions minus two, and above by the smallest marginal proportion. Bounds allow outward rounding uncertainty of 0.05 percentage points for each printed coverage value. They require common denominators and accurate margins but not independent uptake. They do not measure same-day coadministration, dose counts or causal interactions. Both endpoints agree with independent probability linear programs for 105 test cases.

The NDIS district-age panel supplies joint local outcome tabulations, but its service districts and broad age bands do not directly match these PHNs and birth cohorts. At a single receipt cutoff, different-age coverage also reflects eligibility and catch-up time. Without verified geographic allocation, cohort-specific outcomes and a policy counterfactual, these additions receive no causal synthesis weight. Source files, extraction scripts, tests and the source-free reproduction module are linked through source identifier `AUS_NCIRS_COVERAGE_2024` in the evidence register.

Appendix Table S43. Complete event-testing status

Event-specific status	Records	Countries	Share	Interpretation
Outcome analysis linked	414	9	12.3%	Includes limited and zero-weight diagnostics; not an independence count
Attempted; not estimable	12	5	0.4%	Outcome test attempted, but pre/post or outcome support is insufficient
Exposure first stage only	433	157	12.8%	Event-window uptake test exists; no compatible event-specific outcome estimate
Outcome source; timing unverified	215	27	6.4%	Potential outcome source exists; historical date or affected age is not verified
Explicit design gate	12	1	0.4%	Verified supplemental event fails outcome-support or counterfactual requirements
No event-specific outcome estimate	2,288	194	67.7%	Record retained; no compatible repeated outcome estimate in acquired public data

Notes: Statuses are mutually exclusive and sum to 3,379 event records. The 236 strong exposure shifts and 264 related international-analysis matches (120 autism-specific) are cross-cutting audits, not additional event states, event-effect estimates, or independent studies.

This distinction matters for both completeness and validity. A Brazilian municipal MMR-autism gradient, a Finnish exact-cohort MMR panel, or a Swiss canton measles-special-education association is part of the total evidence archive. It is not automatically an estimate of every Brazilian, Finnish, or Swiss MMR-related schedule date. Conversely, a registry row with no compatible repeated outcome is not silently dropped merely because no coefficient can be computed. The event ledger records the failed join. Files 2187-2190 are therefore the authoritative bridge between the frozen 3,379-event global denominator and the broader international analysis archive.

S42. Cohort-step estimates without flexible smoothness adjustment

No-smoothness is not assumption-free. If the unexposed potential-outcome path may jump arbitrarily at the same cohort as schedule eligibility, the causal level effect is not point identified. The estimable objects are instead minimal-adjustment coefficients: exposure-by-post contrasts with unit and cohort or year fixed effects, and national before/after or level-normalized contrasts. Appendix Table S44 reports them even when their direction is unfavorable to the paper's conclusion or their diagnostics fail.

Appendix Table S44. Cohort-step and level estimates without flexible differential trends

Setting/event	Minimal-adjustment contrast	RR	95% CI	Diagnostic status
U.S. IDEA PCV7	Post step x 1-SD state uptake; state/cohort FE	1.017	0.979-1.057	Near one; 51-state intensity contrast
U.S. IDEA RotaShield	Birth-1998 pulse x +10pp uptake; state/cohort FE	1.062	0.978-1.152	Imprecise; one clean pre-exposure cohort
U.S. IDEA RotaShield	Birth-1998/99 window x +10pp uptake; state/cohort FE	1.042	0.964-1.125	Imprecise broad-window sensitivity
SchoolVaxView MMR	PPML, state/cohort FE	0.989	0.930-1.052	Near one; short three-cohort panel
ADDM PCV7	Site and cohort fixed effects	1.072	0.953-1.206	Six-site sensitivity; broad interval
England rotavirus	Fully eligible; age/report-year FE	0.987	0.960-1.014	Near one; national APC-restricted contrast
Scotland rotavirus	Exposure x uptake; authority/year FE	1.195	0.938-1.524	Positive point estimate; imprecise
Wales rotavirus	Exposure x uptake; authority/year FE	1.044	0.822-1.325	Near one; dynamic pretrend fails (p=0.010)
Canada rotavirus	Age-six eligible-cohort TWFE; prevalence	1.153	1.075-1.238	Positive; joint lead p=4.1e-21 (fails)
Pooled schedule entries	First exposed vs cohort -1; no trend extrapolation	1.187	1.109-1.271	Raw rise; visible pre-event ascent
Japan MMR withdrawal	All ASD, pooled post- vs pre-withdrawal	2.031	1.479-2.788	Diagnoses rise after withdrawal; opposite harm sign
Japan MMR coverage	All ASD per +100pp measured coverage	0.249	0.088-0.699	Inverse raw association; secular-trend confounding
U.S. varicella 1995	National post-cohort step; age/year FE	0.870	0.791-0.958	Inverse; placebo-cutoff rank p=0.40

Notes: Rows deliberately remove flexible unit-specific cohort paths or report a raw national level contrast. Effect units remain design-specific. Fixed effects, age alignment, and exposure scaling shown in the contrast column are retained. Without any restriction on the unexposed potential-outcome path, none of the national level contrasts is a point-identified causal effect.

The measured-rollout results are mostly close to one before adding flexible cohort paths: U.S. IDEA PCV7 is RR 1.017 (0.979-1.057), SchoolVaxView MMR is 0.989 (0.930-1.052), ADDM PCV7 is 1.072 (0.953-1.206) per six-site SD (1.345, 0.814-2.224 per ten points), England rotavirus is 0.987 (0.960-1.014), and Wales rotavirus is 1.044 (0.822-1.325). Scotland is positive but imprecise at 1.195 (0.938-1.524). The RotaShield first-exposed pulse is 1.062 (0.978-1.152) per ten uptake points, and its two-cohort window is 1.042 (0.964-1.125). Thus flexible smoothness is not what manufactures the central near-null locations.

The naïve positive rows are also visible. Canada's age-six rotavirus TWFE is RR 1.153 (1.075-1.238), but the joint lead test rejects at p=4.1e-21. In the pooled schedule-entry path, the first exposed cohort is 18.7 percent above cohort -1 (95 percent CI 10.9-27.1 percent) under level normalization. Yet the same raw series has already risen from 22.6 percent below the reference at event time -5 to the reference at -1; the post-event path then keeps climbing to more than twice the reference level. Treating that entire secular ascent as a vaccine effect would attribute the pre-vaccine rise to vaccines as well.

The sign-reversal checks make the problem concrete. Yokohama all-ASD diagnoses are higher after MMR withdrawal, raw RR 2.031 (1.479-2.788), and measured MMR coverage is inversely associated with ASD, RR 0.249 (0.088-0.699) per full-coverage contrast. The national U.S. varicella step is also inverse, RR 0.870 (0.791-0.958), but ranks only p=0.40 against admissible placebo cutoffs. These coefficients are real descriptive results. Their mutually inconsistent signs across additions and withdrawals, failed leads, and sensitivity to trend normalization are why they receive zero causal weight rather than being suppressed.

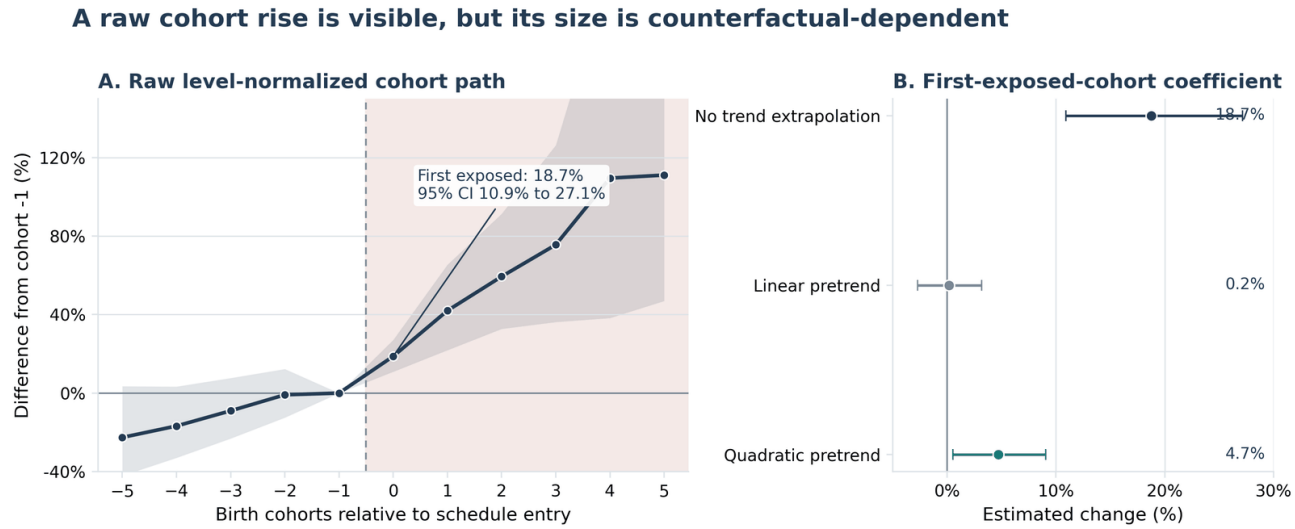


Figure S44. Pooled cohort path without trend extrapolation and first-exposed-cohort sensitivity. The raw level path rises by 18.7 percent at the first exposed cohort, but it is already rising before schedule entry and continues rising afterward. Linear and quadratic pretrend normalizations move the same first-exposed coefficient to 0.2 and 4.7 percent. These are specification diagnostics, not three causal estimates.

S43. Dated acquisition closeout and method-exhaustion audit

S43.1 Stopping rule and source coverage

The acquisition protocol combined a global expansion frozen on 30 August 2026 with a targeted residual holdout closed on 5 September 2026. The dated lock is a stopping rule for the final expansion, not an external preregistration of the earlier project. The versioned `ANALYSIS_PLAN_STATUS.md` distinguishes externally preregistered elements (none), choices frozen before the final locked run, robustness additions, and exploratory or failed designs. “Complete” means that the specified global backbones, country-authority routes, national statistical offices, immunization systems, education and disability systems, public surveys, lower-unit systems, and dated candidate docket all received a recorded disposition. It does not mean that every historical local file, restricted registry, unpublished table, future release, or dataset behind an application has ceased to exist. Access failures, restricted routes, exposure-only files, outcome-only files, and incompatible aggregates are never counted as null results.

The final canonical release contains 212 source-catalog records, 548 evidence-register records, 702 curated policy-event records, 80 subnational source records, and 454 subnational event records. These layers overlap and are not added to the frozen 3,379-event global search frame; the separately reconciled current schedule layer contains 3,422 active canonical records representing 3,391 linked episodes. The final 36-candidate acquisition docket contains eighteen acquired-and-normalized candidates, twelve acquired-raw candidates, three restricted candidates, and three exact technical blocks. Across that docket, 427 acquisition attempts produced 318 successful or reused unique paths, including 174 new files; 168 new files have a valid intended signature and six retained responses are redirects or non-data. The raw protocol archive contains 371 files, 1,648,647,792 bytes, and 370 unique hashes. Its 116 normalized Parquet outputs contain 15,875,489 nonunique rows; this row sum is a processing count, not an analysis sample size.

The regional rescouts were intentionally overlapping checks, not additive jurisdiction counts. One pass reconciled all 108 Americas/Europe backbone entries and 1,761 associated event rows; a second covered 139 Asia, Africa, Oceania, and MENA jurisdictions across five source families each; a third rechecked twenty-seven lower-unit source families across the United States, Canada, the United Kingdom, Australia, and New Zealand. Malaysia's national infant-immunization series was the only genuinely new bulk file in the latter global rescouts, and it is exposure-only. Scripted access blocks in Barbados, the Dominican Republic, Trinidad and Tobago, Malta, Australia, and restricted New Zealand IDI access remain explicit. None is treated as evidence about effects.

S43.2 Marginal panels and event-specific near misses

The newly normalized remaining-public panels yield 104 statistical rows across Mexico, the Philippines, South Africa, and Palestine. All receive zero causal weight. Mexico has no nominal p value below 0.05. Two Philippine acute-malnutrition rows are nominal under two-way inference but not municipality-cluster inference and fail false-discovery-rate correction. One South African COVID-period/broad-disability ratio is nominally positive at RR 1.164 (1.024-1.323) but has q = 0.231, no exposure gradient, and incompatible temporality. Palestine supplies descriptive correlations only; Indonesia is exposure-only. No international row passes the timing, trend, outcome, and counterfactual gates. These rows remain in the archive because failed or descriptive analyses should be visible, not because their count strengthens a conclusion.

An independent adversarial crosswalk adds the Québec campaign and eight India rollout overlays to the 3,379-event core, yielding 3,388 unique event rows. Among the 2,953 unused core rows, in addition to 12 attempted-but-not-estimable rows, 36 have strictly verified timing and 16 have both a material first stage and some outcome asset. None simultaneously has verified timing, affected-cohort support, a material first stage, and a compatible outcome window; none passes all frozen gates. The zero is therefore a documented absence of an unused estimable opportunity in the acquired frame, not a claim that the unestimated events had zero effects.

Québec's 2014 Saguenay-Lac-Saint-Jean MenB campaign is the strongest newly found event/outcome pairing and illustrates why an estimable coefficient is not automatically causal evidence. The official sources provide a treated region, fourteen potential donor regions, and an 83-percent campaign first stage; the primary donor pool retains eleven regions after three exclusions. The ASD outcome is an age-one-to-twenty-four diagnosed-prevalence stock, only two clean post years precede a billing-series break, assignment responded to an outbreak, and the treated pretrend fails. With eleven primary donors, the minimum plus-one rank p value is 1/12 = 0.083. The observed 2015-2016 treated-to-synthetic ratio is 0.869 and the primary in-space p value is 0.083; all 122 Québec rows therefore retain zero causal weight. India supplies official RVV and PCV phase dates plus 57,606 UDID disability-card rows, including 2,451 ASD-coded rows and 122,710 registered ASD cards or certificates, but no public repeated state- or district-by-grade/age/year ASD panel. No India event-study, DiD, synthetic-control, or power estimate is manufactured from that aggregate stock.

The New Zealand Ministry of Education supplies a final outcome-side holdout: age-specific Ongoing Resourcing Scheme (ORS) counts for 2005-2025 and matching age-specific school rolls. ORS is a broad stock of pupils with high or very high ongoing support needs, not autism incidence. For age-five rates, continuous-quadratic cohort-step estimates are RR 0.994 (0.834-1.184) at the 2008 PCV introduction, 0.930 (0.683-1.267) at the 2014 rotavirus introduction, and 0.979 (0.562-1.706) at the 2017 varicella expansion. Their two-sided placebo-cutoff rank p values are 1.000, 0.333, and 0.933. Seventy-two specifications vary age five versus six, total versus high versus very-high need, linear versus quadratic cohort paths, and HC3 versus HAC(1) uncertainty; no HC3 row has p below .05. Leave-one-cohort-out ranges are 0.940-1.023, 0.836-0.981, and 0.836-1.022, respectively, but the last interval endpoints are extremely unstable.

Schedule change	Age-five RR	95% CI	Placebo-rank p	80% MDE
PCV, 2008	0.994	0.834-1.184	1.000	1.27
Rotavirus, 2014	0.930	0.683-1.267	0.333	1.53
Varicella expansion, 2017	0.979	0.562-1.706	0.933	2.02

The table closes a data gap but cannot narrow a causal interval. There is no untreated national counterfactual, the outcome is not autism-specific, and power is poor. The known 2010 ORS eligibility extension is directly visible as 395 extension-category recipients ages five to eighteen. Yet an attempted older-versus-younger internal diagnostic has a rejected pretrend (p = 4.1e-5) and an immediate 2010 relative RR of 1.028 (0.931-1.134), so it fails rather than validates the proposed quasi-experiment. Every New Zealand estimate consequently receives zero causal and synthesis weight.

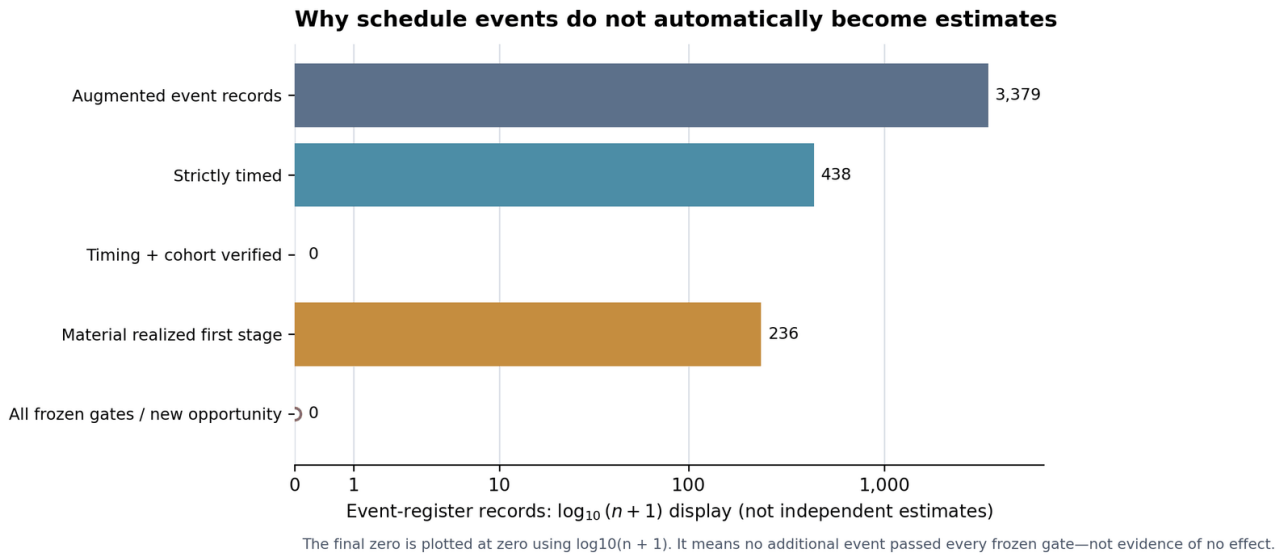


Figure S45. Attrition of the frozen 3,379-event search frame across timing, affected-cohort, realized-first-stage, and full identification gates. Event records are neither independent estimates nor null results; the final zero means no additional unused event passed every gate.

S43.3 Analysis, robustness, and simulation closure

The method-frontier audit covers twenty-four specified design families: cohort and spline trends; synthetic control, synthetic DiD, and matrix completion; modern staggered-adoption estimators; randomization, permutation, wild-cluster, serial, and spatial inference; HonestDiD; generated-exposure and suppression propagation; placebos and positive controls; interactions, removals, age and latency; IV and shift-share diagnostics; Bayesian synthesis; and the individual observational calibration. The audit identified no additional estimator within these families that resolved the documented support failures in the held data. The binding limitations are common national timing, absent untreated donors, too few preperiods or clusters, failed pretrends, no measured local first stage, broad or changing outcome definitions, and missing fixed-age cohort outcomes. Applying another estimator after one of those failures would create another number, not another identified result.

Dependence and multiplicity were re-audited across all 25,597 empirical and descriptive disposition rows. The three canonical strict-likelihood rows are not nominally significant, and their estimates and posterior do not change. Several apparent positives disappear under their required inference: state-MMR p = 0.000035 becomes 0.355 with state-and-year clustering and 0.118 under wild inference; the two Philippine rows have q = 0.677; and the South African row has q = 0.231. A joint eight-outcome PCV test rejects an undirected all-zero vector, but the maximum positive coefficient has familywise p = 0.606. Thirty-two synthetic-control/placebo rows have calibrated rank inference and zero causal weight; no donor-supported hidden rejection is present.

Power is evaluated on each design's native contrast. Exact effect-injection experiments refit the PCV7 and plug-in RotaShield pipelines after injecting null and positive effects. PCV7 has 86.8-percent recovery at RR 1.05 and an 80-percent MDE of about 1.047. A separate calibrated variance calculation propagating 249 RotaShield exposure redraws under the four-test Bonferroni rule gives 83.3-percent power at RR 1.10 and an 80-percent MDE of about 1.096; it is not a joint exposure-and-outcome injection experiment. Simulation and calibrated variance calculations describe performance under their stated data-generating and uncertainty models; they are not empirical evidence about vaccination.

An independent operating-characteristic suite separates plug-in estimator behavior from exposure-uncertainty propagation. Repeatedly refitting the exact 353-cell PCV7 estimator under independent Poisson sampling gives 5.2-percent null rejection, 94.8-percent interval coverage, and 86.4-percent power at RR 1.05. State-level AR(1) shocks produce 6.0-percent null rejection, whereas shared Census-division AR(1) shocks combined with state-only clustering raise it to 11.2 percent. A dedicated nine-division audit reproduces RR 1.0128, gives CR1 RR 0.976-1.051 (p = 0.447), and gives CR2/Satterthwaite RR 0.974-1.053 with 5.15 degrees of freedom (p = 0.442). Exhaustive enumeration of all 512 null-imposed division sign patterns gives p = 0.531 for RR 1, p = 0.070 for RR 1.05, and p = 0.0078 for RR 1.10; leave-one-division-out estimates range from 0.999 to 1.021. The two-lead PCV7 pretrend test detects only 20.5 percent of one-percent annual differential drift and 65.9 percent of two-percent drift. A treatment-coincident confounder is, by

construction, indistinguishable from a causal jump. These diagnostics delimit the identifying assumption rather than convert it into a testable fact.

The same suite clarifies two other finite-sample limits. For RotaShield, the plug-in wild-residual procedure has 94.3-percent power at RR 1.10 in 1,000 effect draws (1,500 null holdout draws), while the separate 249-redraw generated-exposure calibrated variance calculation reported above gives 83.3-percent power; the two percentages answer different uncertainty questions. The nine-equation local-MMR sign-flip test has 4.6-percent null rejection under independent symmetric errors but 42.1 percent under a strong common shock and only 40.7-percent power against a common RR 1.02. It is therefore a finite-sample diagnostic, not the principal small-effect test. Synthetic-control rank tests are also resolution-limited: with three donors their minimum attainable p value is 0.25, and with fourteen it is 0.067. The strict Bayesian stress test discriminates a common five- or ten-percent generalized effect when it reaches the observed contrasts, but a fully masked common effect is observationally equivalent to the null. No simulation can remove that limitation.

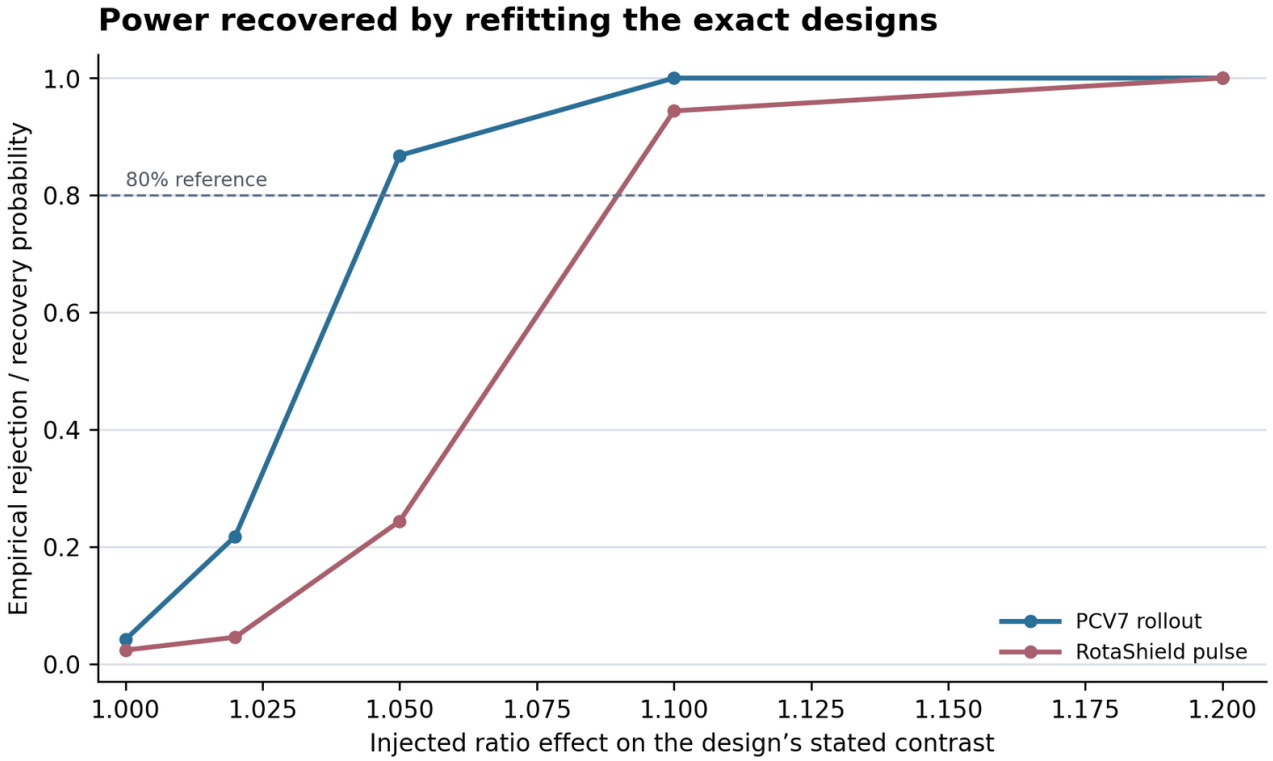


Figure S46. Empirical rejection and recovery rates from exact-pipeline effect-injection experiments for the PCV7 rollout and RotaShield pulse. Curves are design-conditional simulation results on native contrasts, not another empirical effect estimate.

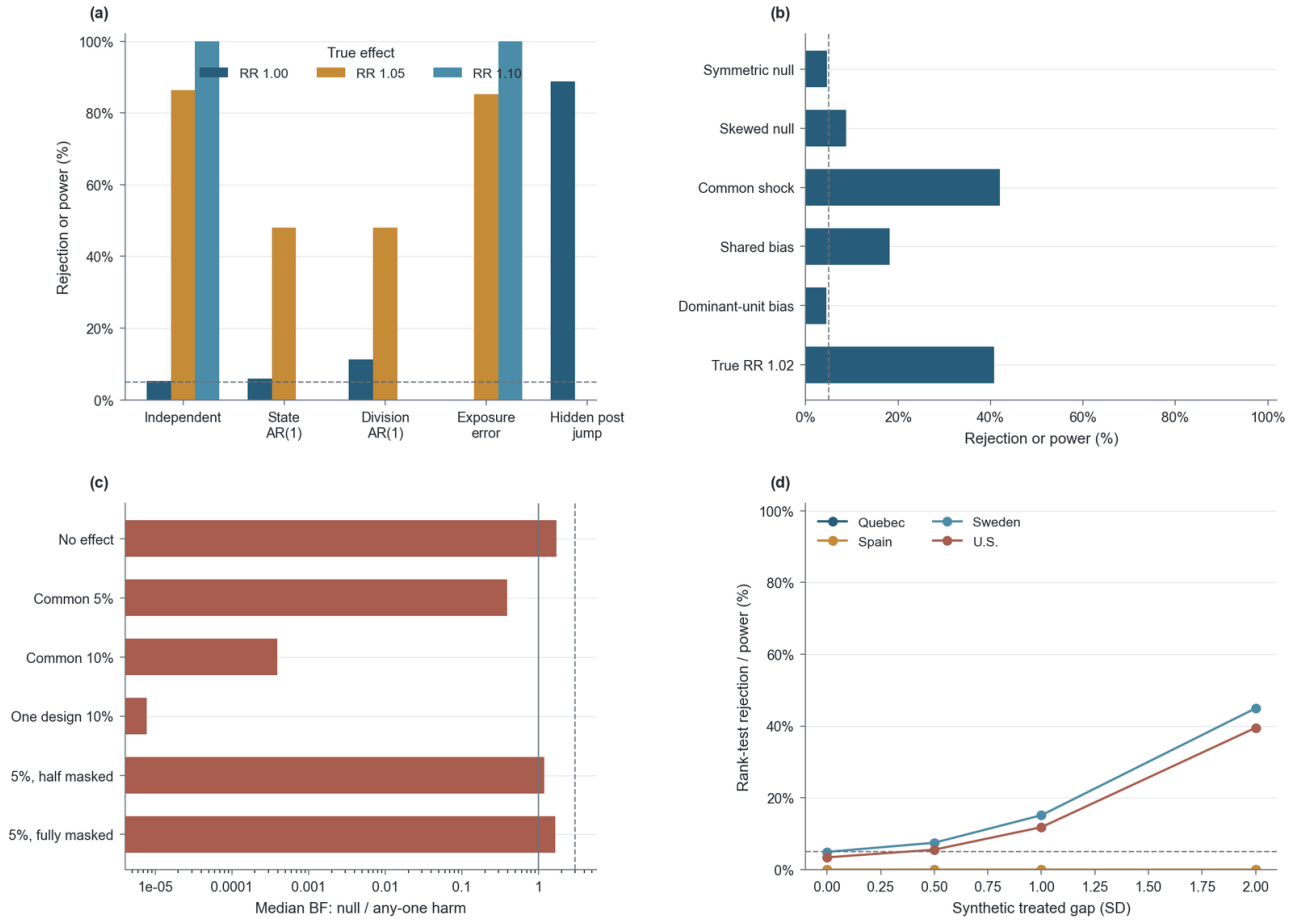


Figure S47. Synthetic operating-characteristic stress tests for the PCV7 estimator, local-MMR sign-flip inference, strict-design Bayesian prediction test, and finite-donor synthetic controls. The panels diagnose behavior under stated simulated violations; they are not empirical vaccine-effect estimates.

The Japan withdrawal is also reported as a complete specification and masking audit rather than a selected coefficient. The continuous-exposure specification curve shows how the estimate changes as cohort curvature is relaxed. In the structural display, the illustrative predicted-decline path holds the unvaccinated background diagnosis rate at its 1988-implied value; it is not an estimated no-vaccine counterfactual. The display then shows the additional acceleration in that background path required to conceal candidate harmful effects. These are transparent assumption stress tests. They do not repair the near-collinearity between declining MMR coverage and cohort time, do not promote the withdrawal to a causal pool, and do not allow its theory-design weight in the prediction stress test to narrow a native causal interval.

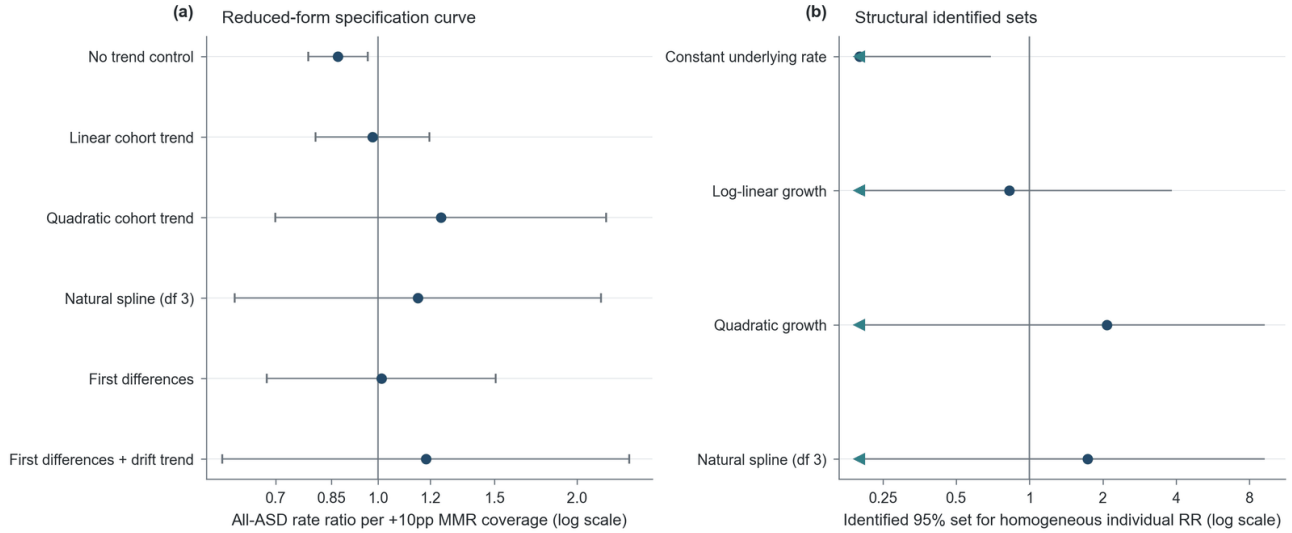


Figure S48. Japan MMR-withdrawal specification curve. Raw, linear-trend, difference, curvature, spline, and drift-sensitive estimates are shown without selecting a preferred favorable coefficient; coverage and cohort time are nearly collinear.

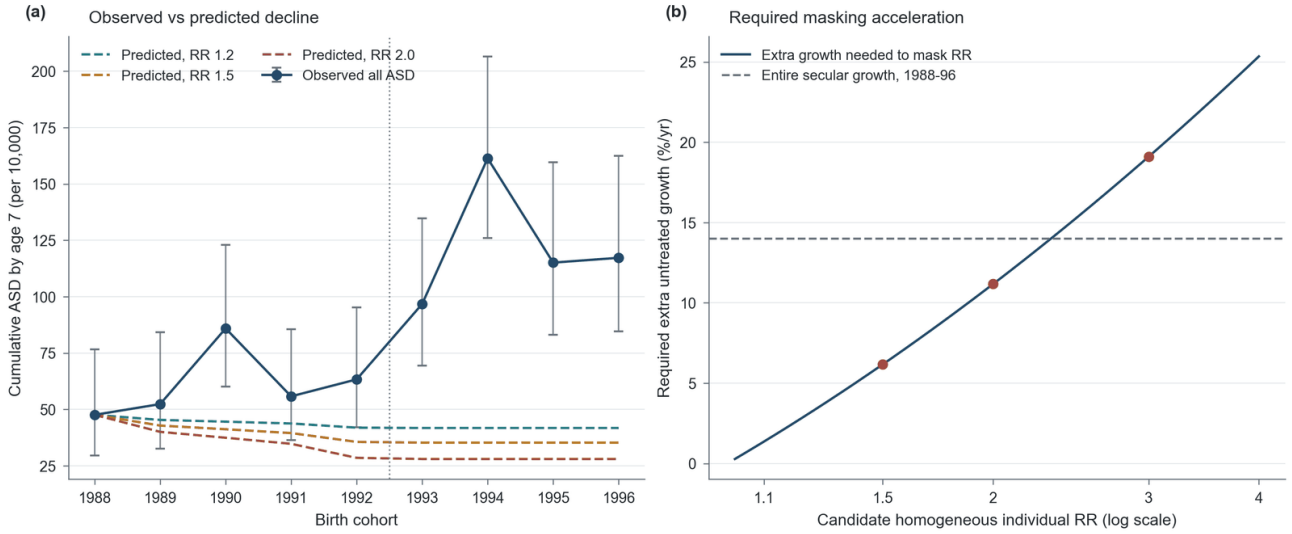


Figure S49. Japan withdrawal masking audit. Candidate harmful effects imply cohort declines after exposure collapses; the display quantifies the countervailing unvaccinated-outcome acceleration required to conceal those declines under the stated homogeneous mapping.

S43.4 Descriptive completeness and conclusion impact

The archive inventories 214 candidate result files, of which 191 contribute 25,559 retained statistical rows. Among 24,537 ratio-scale candidates, 24,243 supply auditable ratio likelihoods across 175 source files and reduce to 23,074 distinct numerical signatures. A separate ledger retains 683 auditable additive or otherwise nonratio native-unit likelihood rows without forcing them onto an RR scale, including all 156 Korean additive rows paired with the retained Korea ratio estimates. The broader 25,597-row disposition ledger starts from those 25,559 rows, removes 89 embedded panel rows, and adds 104 current-panel rows plus 23 individual-calibration rows. None receives positive weight merely by entering an accounting catalog. Source-specific fail-closed extraction admits seventy-one valid Swedish ratio rows, explicitly excludes one nonpositive interval, and admits no malformed ratio interval. All retain zero synthesis weight, so this broader accounting does not alter the primary inference.

The closeout joins 9,437 exposure rows to 13,646 outcome rows and finds 208 exact geography-by-cohort support pairs; 204 have at least three contiguous overlapping cohorts. Those are support diagnostics, not 208 independent causal designs. The dependence audit maps the current evidence register into connected components rather than counting overlapping records as independent. Thirty-seven synthetic-control design spines are inventoried and all retain zero causal weight. Three indexed result files remain absent from the native likelihood ledger, but none contains a valid missing ratio likelihood. Across the 3,379-event frozen search frame, sequential timing, affected-cohort, first-stage,

outcome, counterfactual, and preperiod gates leave zero additional unused event passing every gate. The zero records a design frontier, not a null effect for the failed events.

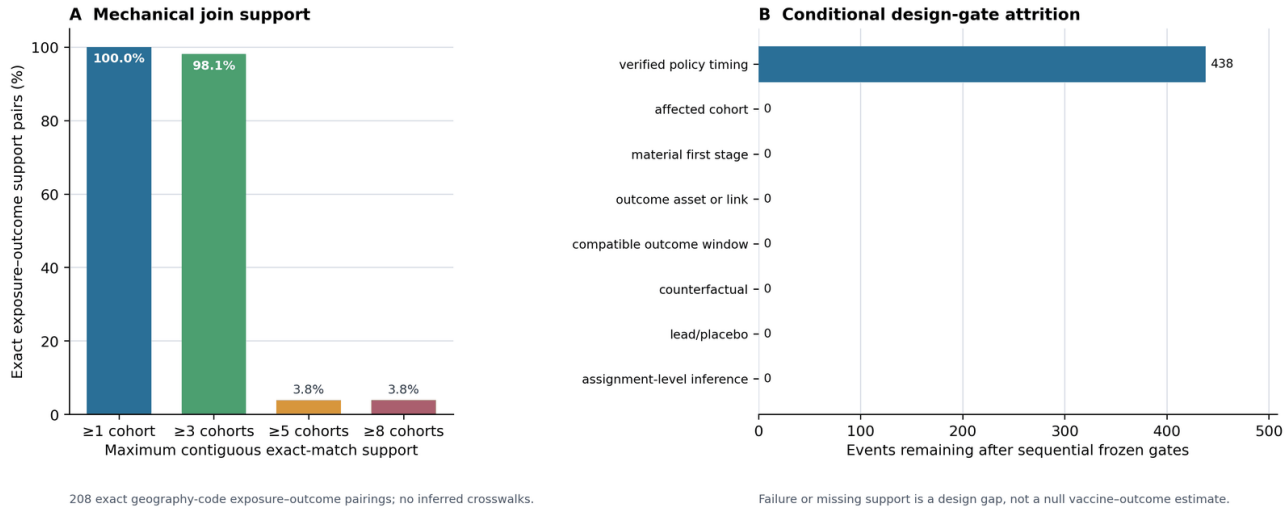


Figure S50. Descriptive closeout of exposure-outcome support and sequential design gates. Counts describe available joins and diagnostics, not independent effects; the final zero means that no additional unused event in the frozen search frame passed every causal gate.

The supported policy contrasts provide the causally informative layer. The observational literature supplies separate triangulation, while the three-design calculation remains a scale-dependent prediction stress test rather than a causal synthesis. Component sources, access dispositions, and zero-weight diagnostics do not change a causal likelihood. The search frame and source-specific stopping rules define the scope of this inventory; they cannot establish the absence of every other public dataset.

S44. Source reconciliation and uncertainty in individual-risk calibration

Numerical and source reconciliation here verifies computations conditional on a stipulated mixture equation. It does not validate the equation's application to the PCV7 spline coefficient; that application fails the exposure-aligned audit in Section S52. PCV7 posterior rows below are algorithmic sensitivities, not individual-risk evidence.

S44.1 County and local-authority outcome measurements

The CDC pilot dataset 7vg3-e5u2 links statewide health and education records for 2018 residents ages three through twenty-one. It reports 394 counties across Arizona, Arkansas, Minnesota, New Jersey, Tennessee, Utah, and Wisconsin; nine county case counts are suppressed. Observed-county prevalence uses only the population of counties with reported counts. Dividing reported cases by the population of all counties instead supplies a lower bound, and is labeled accordingly. No suppression-based upper bound is calculated without a documented suppression threshold.

State	Observed-county ASD prevalence	Whole-state lower bound
Arizona	2.127%	2.127%
Arkansas	1.686%	1.683%
Minnesota	1.570%	1.567%
New Jersey	1.845%	1.845%
Tennessee	1.513%	1.509%
Utah	2.066%	2.064%
Wisconsin	1.478%	1.477%

These are administrative prevalence measurements, not vaccine-effect estimates. The broad age range, single outcome year, and potential person overlap with education records preclude treating the county rows as independent policy studies. The data and metadata endpoints are https://data.cdc.gov/resource/7vg3-e5u2.csv and https://data.cdc.gov/api/views/7vg3-e5u2.json. Script 3001 downloads, verifies, and summarizes the source while preserving missingness.

Emilia-Romagna's official 2023 autism-service factsheet reports eight local health authorities (AUSLs), with separate child and adult counts. The child AUSL counts sum to 6,575, whereas the regional distinct-person count is 6,558; the source explicitly permits a person to be counted by multiple AUSLs. Those 17 excess memberships must not become extra people or independent observations. The adult total is 1,194. The regional childhood service count rises from 1,584 in 2011 to 6,558 in 2023, but these broad-age service endpoints have neither a matched exposure contrast nor an untreated counterfactual. Script 3003 records the source URL and hash, extracts the sixteen local count cells, and checks the overlap. This outcome component helps audit measurement and coverage of Italian sources; it supplies no vaccine coefficient.

The public Socialstyrelsen ADHD/autism dashboard exposes 3,645 ADHD-care indicator rows for 2015-2023 across 27 national, regional, and area-type groups. Its three indicators concern melatonin treatment, stimulant medication after incident ADHD, and transition from child to adult care. The exported semantic table contains no autism diagnosis indicator. Script 3008 retrieves the public dashboard metadata, chart definitions, and data; script 3007 verifies the archived dimensions. This descriptive panel is not an autism outcome, a vaccine-effect estimate, or an automatic negative control. The held Swedish F84 county-age-year series remains the autism outcome source.

S44.2 Japanese coverage reconciliation

Japan's MHLW municipal MR tables for fiscal years 2024 and 2025 contain 1,737 municipalities and 47 prefectural totals per dose phase and year. Prefectural totals are checked against their constituent municipalities and then excluded from municipal sums. Measles coverage includes MR and monovalent measles doses. Weighted municipal coverage rises from 92.652% to 96.248% in phase one and from 91.019% to 92.674% in phase two. Administrative rates above 100% are retained and flagged, rather than silently capped. They indicate numerator-denominator discrepancies that require care in any exposure model.

The fiscal-year files are already represented in the source archive and do not constitute independent new datasets. The matched comparison and sum checks are generated by script 3004. Similarly, the Swedish municipality/county PxWeb retrieval reproduces the held coverage series and adds no independent observations. Neither exposure-only archive supplies a new autism outcome or increases causal precision.

S44.3 Joint uncertainty in the ecological-to-individual map

The archived script 3006 jointly redraws survey shares and refits ecological coefficients using a Beta/Kish approximation. Preserving coefficient-share dependence is useful, but it does not repair the unsupported substitution of a raw coverage-pair mixture for the PCV7 spline estimand. Its individual-risk posterior table and figure are not current inferential results.

Exposure uncertainty for a valid individual-scale model must recompute the relevant cohort receipt path and estimator together, respecting available survey design, cross-wave dependencies and shared exposures. A complex-survey bootstrap, a justified joint measurement model, or an uncertainty-set analysis can be appropriate. Catch-up scenarios and product-definition alternatives are structural sensitivities, not sampling draws with known probabilities. The S53 mixture estimates currently hold reconstructed exposure fixed and therefore do not claim that this task is complete.

S45. Age-group comparisons and counterfactual-trend sensitivity

Comparing policy-eligible children with older, policy-ineligible age groups can remove common calendar shocks without requiring individual vaccination records. The comparison identifies a policy effect only if age-specific counterfactual trends are credible and the comparison group is unaffected. Older groups are not necessarily unvaccinated, and differential changes in diagnosis or access to services need not cancel. With a log-rate outcome the contrast would be a ratio of rate ratios, not automatically an individual vaccinated-versus-unvaccinated risk ratio. Reusing the same records across ages, vintages or estimators does not create independent evidence.

An age comparison can estimate the same policy contrast as another design when its target population, outcome age, exposure contrast and follow-up are held fixed. That equivalence is not automatic: changing outcome ages or comparing earlier with later doses changes the target contrast, and effects in the comparison group are subtracted from effects in the target group. Birth cohorts must therefore be tracked as they move through age bands. A group that is older at the outcome date is not necessarily ineligible at the historical vaccination date. Common calendar effects can be removed without assuming a flat age gap, but extrapolating a nonflat gap still restricts the counterfactual. Neither more age cells nor more estimators resolves unrestricted age-period-cohort confounding.

Québec's province-wide SISMACQ export supplies 675 age-by-sex-by-year observations across autism and ADHD, for fiscal-year starts 2000–2024. Autism is cumulative diagnosed lifetime prevalence; ADHD is a separate substantive outcome rather than a negative control. The source indicator codes are 0303 and 0302, respectively. Publicly funded infant rotavirus vaccination began on 1 November 2011, with doses scheduled at two and four months (Direction régionale de santé publique de Montréal, “Rotavirus,” https://santepubliquemontreal.ca/professionnels-et-partenaires/thematiques-de-sante-publique/rotavirus).

The exploratory age-gap regressions use a ramp in the difference between the estimated fractions of policy-eligible birth cohorts in the younger and older age bands. Uniform births, equal cohort sizes within each age band, an annual observation midpoint and alternative birth cutoffs are explicit approximations, not measured uptake. Exact historical eligibility and fiscal-age conventions, overlapping schedule changes and the 2016 billing discontinuity remain identification limits. The coefficients are percentage-point changes per full-unit eligibility contrast, not individual relative risks. Regional fine-age cross-tabs are absent from this export.

The sensitivity grid attempts 3,888 specifications, of which 2,268 have sufficient rank and support. It varies sex, younger and comparison ages, follow-up endpoint, cutoff, constant/linear/quadratic gap trends, differential billing and pandemic terms, and autocorrelation lags. All fitted specifications remain exploratory and receive zero causal-synthesis weight. Repeated specifications are not independent tests of the vaccine hypothesis.

Diagnostic	Result	Interpretation
HAC null rejection in specified simulation scenarios	13.6–45.3% at nominal 5%	Conventional short-series intervals are inadequately calibrated
Fixed-covariance GLS null rejection	3.95–6.45% at nominal 5%	Calibration holds approximately under the stipulated Gaussian AR(1) model
Autism conditional 80%-power threshold	0.191–0.747 percentage points	Depends on comparison ages, assumed autocorrelation and fitted noise; not an individual-RR exclusion bound
ADHD conditional 80%-power threshold	0.631–4.021 percentage points	These comparisons cannot reliably resolve small changes
Prepolicy forecast checks	432	Probe counterfactual extrapolation; not independent causal replications

The simulation archive contains 96,000 HAC size/power draws and 80,000 fixed-covariance GLS calibration draws. GLS intervals condition on five specified autocorrelations rather than claiming that the correct covariance is known. Neither simulation calibration nor a flexible trend repairs confounding. Published rates are rounded to 0.1 percentage points; worst-case linear coefficient perturbations from rounding are reported separately from statistical intervals, without introducing pseudocounts or invented denominators.

A circular residual-block bootstrap supplies a further serial-dependence sensitivity check, preserving fixed regressors and resampling centered OLS residuals with a residual-degrees-of-freedom correction. Thirty-two basic 95% intervals cover eight age contrasts and block lengths of two through five years, using 999 resamples per interval. Conditional Gaussian simulations evaluate 9,600 series at four block lengths, with 400 simulations per

design/covariance/block-length cell. At nominal 5%, null rejection is 7.75–21.25% with independent errors, 14.5–32% with AR(1) correlation 0.5, and 21.25–45% with correlation 0.8. Monte Carlo standard errors are at most 2.5 percentage points. These basic intervals are not inverted null-imposed tests; their calibration failures preclude treating negative coefficients as protection or their bounds as reliable exclusion limits. Resampling does not repair uncertain exposure assignment or diagnostic confounding. The method draws on the fixed-regressor block-bootstrap framework (Block Bootstraps for Time Series With Fixed Regressors); the finite-sample diagnostics, not asymptotic validity alone, govern its interpretation here.

Pre-only constant, linear, quadratic and three-knot natural-spline forecasts provide an additional 96 early-policy diagnostics. Splines extrapolate linearly beyond their last preperiod knot. For autism at ages one to four versus eighteen to twenty-four, the mean 2011–2013 forecast gap is -0.115 percentage points under a linear trend, +0.024 under a quadratic trend and -0.019 under the spline. These are forecast errors, not cohort-eligibility effects; their sign sensitivity and rounding uncertainty prevent a protective interpretation.

Federal CCDSS exports provide fine age bands nationally, but only pooled ages one to nineteen provincially. Six refreshed export views contain unchanged data cells and no provincial fine-age observations. The Québec component is not independent of Québec administrative records. CCDSS reports that a reliable adult autism case definition was unavailable (Public Health Agency of Canada, “Understanding autism data,” https://health-infobase.canada.ca/autism/).

S45.1 Measured uptake and concurrent vaccination changes in Quebec

Original INSPQ coverage surveys distinguish policy eligibility from vaccine receipt. Rotavirus completion by fifteen months was 13.6% in the 2012 survey, 85.9% in 2014 and 85.8% in 2016; corresponding completion by twenty-four months was 12.5%, 78.2% and 86.9%. Survey year is not birth year: the younger samples were born July-September 2010, 2012 and 2014, respectively. These observations demonstrate substantial uptake change, but not a zero-to-universal first stage. They cannot be assigned indiscriminately to every age-band member. Sources are INSPQ reports 1651, 1973 and 2341, archived with hashes and birth-window checks in the analysis directory.

Concurrent changes complicate an isolated rotavirus interpretation. Varicella vaccination moved later in infancy, while hepatitis B coverage subsequently expanded. The November 2011-June 2012 school measles campaign also means older children, and adults who were schoolchildren at the time, are not necessarily unaffected comparisons. The age-gap analysis therefore includes 1,120 omitted-campaign-effect sensitivity scenarios and 160 joint diagnostics. For an assumed positive 0.1-percentage-point effect among school-eligible cohorts, the modeled omitted-variable bias in the autism age-one-to-four contrast ranges from -0.081 to +0.024 percentage points across specified timing and covariance assumptions. This is a hypothetical bias calculation, not an estimate that the campaign affected autism. Algebraic recovery checks verify the sensitivity calculation; they do not validate the assumed school birth windows or resolve coincident interventions.

S45.2 Earlier second-MMR eligibility and Danish hospital-use age contrasts

Denmark moved routine MMR2 from age twelve to four on 1 April 2008; older children retained the age-twelve schedule. This changed second-dose timing, not first MMR exposure. SSI reimbursement-based reports show delayed receipt and distinguish whole-cohort from eligible-cohort denominators: the 2010 bulletin reports 57% for the whole 2004 cohort and 64% for its measured eligible portion. Later reports warn that some figures include the quarter-cohort not offered the early dose. These figures cannot supply a harmonized individual first stage (SSI, EPI-NYT 10/2010, 21/2011, 23b/2012 and 16/2013).

INDAMP03 supplies 6,630 population-rate cells for 2006-2018: five somatic hospital-use measures, two overlapping diagnosis aggregates, three sex groups and ages one to seventeen. These are not autism or disability prevalence. The 3,240 diagnostic fits compare ages six, eight or ten with two or three, varying constant/linear log-gap trends, three birth-year alignments and three fixed autocorrelations. Historical age and denominator definitions remain unresolved; all fits have zero causal-synthesis weight.

Older/comparison ages	Hospital-patient ratio of rate ratios	Conditional 95% interval
6 / 2	0.942	0.829-1.071
6 / 3	0.955	0.878-1.038
8 / 2	0.975	0.815-1.165
8 / 3	0.991	0.844-1.162

Older/comparison ages	Hospital-patient ratio of rate ratios	Conditional 95% interval
10 / 2	1.077	0.883-1.315
10 / 3	1.103	0.962-1.266

This reference slice uses all diagnoses and sexes, a linear age-gap trend, year-end birth-year alignment and assumed AR(1) correlation 0.5. The ratios describe hospital-patient rate contrasts associated with earlier-dose eligibility, not individual vaccine RRs. Intervals spanning one are not equivalence tests. All rows reuse the same hospital system and cannot be pooled as independent studies.

A separate 360-scenario simulation uses 600,000 Gaussian AR noise vectors, reused across fitted-covariance and injected-effect assumptions. With correctly specified covariance, nominal 5% rejection ranges from 4.765% to 5.235%; allowing covariance mismatch expands the range to 1.49% to 28.45%. Correct-covariance power against a 10% age-gap rate contrast ranges approximately from 16% to 71%. These calculations condition on a linear mean, year-end eligibility and noise estimated from the observed series. They show why fixed-correlation intervals cannot be presented as unconditionally reliable or as excluding small individual effects. Exact noise-free synthetic recovery and rounding-bound checks establish computational properties, not identification.

Pre-eligibility forecast validation fits constant, linear and quadratic log age-gap trends using only training-period outcomes, then holds out the final one or two years before the first affected cohort reaches the outcome age. The full grid contains 3,240 comparisons across all five measures, diagnosis aggregates, sexes, age pairs and alignment scenarios. Of these, 2,160 have at least four training years and two residual degrees of freedom; 1,080 remain explicitly unsupported. For all-diagnosis hospital-patient rates, both sexes, year-end alignment and a one-year holdout, the twelve supported age-eight/ten comparisons give observed-to-forecast age ratios of 0.893-1.118 across the three trend degrees. These are forecast discrepancies, not vaccine effects or confidence bounds. Some validation years follow the calendar policy date despite preceding direct cohort eligibility, so spillovers and concurrent changes remain possible. Nearest-integer rounding scenarios are separate from sampling uncertainty. Independent parsing and polynomial reconstruction reproduce every supported calculation. The comparisons reuse the same observations and have zero causal-synthesis weight.

Scripts, results and original coverage vintages: qa/completeness_20260907/age_control_designs. These diagnostics neither alter the principal causal synthesis nor establish that vaccination reduces autism.

S45.3 Brazilian child-adult hospital comparisons

The SIH/SUS extension covers all twenty-seven processing states in 2003-2007, using 1,619 available monthly archives. October 2007 for Amapa is absent from the official directory. We retain that gap and compare an eleven-month window excluding October in every state and year with all available months. Counts are normal authorizations with valid dates and ages below sixty-five, not unique patients. The principal analysis here restricts admission and processing to the same year; authorizations processed later are not thereby shown to be erroneous or irrelevant. Resident-state numerators use July-1 population estimates, not the hospital processing state's population. The new and older child pipelines reconcile exactly across 135 state-years and 12,717,231 child authorizations; this is computational verification of overlapping data, not independent replication.

Contemporaneous Santos notices report a local rotavirus launch on 6 March 2006, with the first dose limited to young infants and the second conditional on prior receipt. This is not proof of identical municipal implementation dates. Some late-2005 births could qualify, while age-zero outcomes in 2006 include prelaunch and prevaccination time. Ages one to four mix eligible and ineligible cohorts. Neither the calendar introduction nor older comparison ages establish individual vaccination status or absence of indirect effects (Prefeitura de Santos, “SMS inicia vacinacao contra o rotavirus,” 6 March 2006, https://www.santos.sp.gov.br/?q=noticia/sms-inicia-vacinacao-contra-o-rotavirus).

The Ministry of Health's historical bulletin corroborates national implementation beginning March 2006, but does not establish identical local dates or individual receipt. It also describes 2004 child and adult measles-containing vaccine recommendations. Older ages are therefore ineligible for the infant rotavirus addition, not necessarily unaffected by other schedule changes. These cointerventions remain unresolved in the age-gap diagnostics (Ministerio da Saude, Boletim Epidemiologico 46(30), 2015, p. 2, https://www.gov.br/saude/pt-br/vacinacao/publicacoes/besvs-pni-v46-n30.pdf).

The archive retains 864 unadjusted age contrasts across six recorded endpoints, comparison ages, calendar windows and coded-age-disagreement sensitivities. A separate state-fixed-effect model fits seventy-two complete log-age-gap

profiles, retaining 144 zero-rate support failures rather than dropping states or adding pseudocounts. The table shows the equal-state mean log-rate contrast for gastrointestinal principal diagnoses (A00-A09), infants versus ages twenty-five to forty-four, with 2005 as reference and October excluded throughout. It is not the same weighting as the pooled national contrast.

Year	Ratio of rate ratios	State-cluster conditional 95% interval	Region-cluster conditional 95% interval
2003	1.013	0.937-1.095	0.897-1.144
2004	1.036	0.970-1.107	0.950-1.131
2006	0.935	0.857-1.019	0.869-1.005
2007	0.701	0.637-0.771	0.616-0.798

The two-preperiod joint tests give conditional p values of 0.546 with state clustering and 0.413 with five-region clustering; nonrejection does not establish parallel trends. Regional sign-score enumeration gives p = 0.0625 for 2007, the minimum attainable two-sided value with five regions. Its validity requires independent regional scores symmetric about zero, not merely an observed decline. In 60,000 simulated Gaussian noise panels, shared regional shocks raise nominal 5% state-clustered rejection to 25.56-47.16%, whereas region clustering gives 4.48-4.96% in those scenarios. A common national age-specific shock undermines both procedures. These results favor reporting dependence assumptions explicitly, not choosing whichever interval excludes one.

The joint pretrend test needs a separate calibration. Across 60,000 additional noise panels and 168 test/drift scenarios, its five-region nominal 5% rejection rate is 8.96-11.02% under independent-state or regional shocks, and 81.26-82.18% with half the variance assigned to a national shock. Thus the preceding single-coefficient calibration cannot validate the joint test. A scenario-specific null threshold estimated from 2,500 draws and evaluated on 2,500 separate draws gives only 8.48-24.80% detection of a 5% annual age-gap drift under entirely regional shocks; detection of a 10% drift is 18.20-64.88%. The corresponding held-out null rejection is 4.60-5.52%. These are conditional simulation benchmarks, not a feasible correction with known real-world covariance, and the injected drift is not an individual vaccine-risk ratio. The observed pretrend nonrejection is therefore weak reassurance. Even a powerful pretrend check cannot exclude confounders that begin only after introduction.

Rare principal diagnoses remain weak measures of the claimed outcomes. Only 406 same-year F84 records occur in the selected national data; these are not population autism incidence. A zero-compatible, independent-Poisson conditional benchmark for infants versus ages twenty-five to forty-four in 2007 relative to 2005 uses counts [[0,1],[11,24]], giving a boundary rate-ratio contrast of zero with interval 0-95.19. The corresponding intussusception benchmark is approximately 1.24 (0.77-2.02). Neither supplies an individual vaccine-risk bound: repeated authorizations, ascertainment, exposure alignment and cross-unit shocks are not resolved by an exact count interval. All 432 rare-count benchmark specifications remain separate from the causal synthesis.

Finally, 18,140 blank/unknown-principal-diagnosis records have invalid residence prefix 00 and cluster in February 2006. Their exclusion from resident rates must not be displayed as evidence of zero unknown diagnoses. Independent fixed-width reading confirms this source pattern in the Sao Paulo and Parana files, without identifying its administrative cause. Complete counts, source hashes, alternative contrasts, native-rate and percentage-change SVGs, simulations and validation scripts are retained in the age-control analysis directory. These extensions add descriptive and design-diagnostic information, not an autism-null replication or a new independent causal likelihood.

S45.4 Brazilian birth-cohort composition and threshold sensitivity

The longer child panel allows comparisons at the same completed age, but in different calendar years. Across three birth-cohort comparisons, eighteen ages and three principal-diagnosis outcomes (F840/F841, G40/G41 and G80), 138 of 162 attempted contrasts have complete observation windows; 131 have a defined diagnosis-share ratio. The denominator here is all child hospital authorizations, not resident population. A separate residence-state aggregation matches all 552 numerator and denominator comparisons. This verifies computation on shared records, not independent epidemiological replication.

The F840/F841 share ratio for births in 2007 versus 2004 is 10.664 at age five and 0.274 at age ten. The latter becomes 1.392 when Sao Paulo is omitted; 36 of the earlier cohort's 42 authorizations at that age are from that state. All 3,726 leave-one-state-out calculations are retained. Geographic concentration is not proof of coding error, and neither the increases nor decreases establish a vaccine effect.

Using registered births instead of all hospital authorizations as the denominator provides a complementary sensitivity check. The SINASC archives contain 3,026,548 registered births in 2004 and 2,891,328 in 2007. At age five, the three versus twenty-eight F840/F841 authorizations correspond to 0.099 versus 0.968 authorizations per 100,000 births, a ratio of 9.770. At age ten, forty-two versus eleven authorizations correspond to 1.388 versus 0.380 per 100,000 births, a ratio of 0.274. Thus replacing the hospital-share denominator does not eliminate this contrasting age pattern. These are authorizations relative to original cohort births, not diagnosed individuals, surviving resident populations or person-time rates. Repeated admissions, migration, mortality, care access and diagnosis practices remain relevant. The cohorts reach each age in different calendar periods, and denominator sensitivity is not independent evidence of a vaccine effect. Source hashes, observation-window gates and the full comparison table are retained with the replication scripts.

Among the pre-eligibility birth cohorts, the defined F840/F841 authorization-per-birth ratios for 2004 versus 2003 range from 0.531 to 4.015 across ages. The largest ratio uses only one versus four records. This is a comparison of cohorts born before routine infant eligibility, not a clean pre-policy-period parallel-trends test: most outcome windows occur after rollout, and indirect policy exposure is not excluded. The variation illustrates why sparse cohort ratios alone cannot distinguish vaccine effects from other age, calendar and cohort differences.

Age-period Poisson models instead compare calendar cells using age and year fixed effects, with optional age-specific linear or quadratic trends and an offset for all hospital authorizations. The principal cohort contrast separates births from 2007 onward from births through 2004, excluding transitional cohorts. All eighteen F840/F841 specifications yield positive conditional share ratios, ranging from 1.488 to 2.390. These are not null results. A hospital denominator that can itself change with vaccination or service use, unresolved cohort-specific diagnosis production, unmeasured receipt and uncalibrated dependence prevent interpreting them as individual vaccine-risk estimates. Saturated cohort effects absorb the exposure indicator exactly; full rank under restricted age-period controls does not validate those restrictions.

First later birth cohort	F840/F841 conditional share-ratio range
2003	0.993-2.986
2004	1.616-3.261
2007	1.488-2.390
2010	0.273-0.864
2012	0.138-0.627

Each threshold omits the two preceding birth years and therefore changes cohort support. These are threshold sensitivities using later outcomes, not clean prepolicy placebos or confidence intervals. Across all three outcomes, 270 conditional models and ninety deterministic mean-recovery checks are retained. Calendar-input reconciliation matches 3,198 cells and checks 258 implicit zeros against separate residence-state aggregates. The changing signs warrant further study of cohort and ascertainment patterns; they do not by themselves prove either confounding or vaccine causation. None of these overlapping diagnostics contributes an independent causal likelihood.

To isolate denominator sensitivity from sample composition, a separate matched-cell analysis restricts both offset models to birth cohorts 1996-2017, retaining the 2007 cutoff and excluding births in 2005-2006. It estimates eighteen specifications per outcome and denominator: three age windows (1-17, 4-12 and 5-17), two end years (2019 and 2025), and age-specific trend degrees zero, one and two. Source years 1999, 2000 and 2007 are excluded for archive completeness; the longer window also excludes 2020-2022. Each denominator pair uses identical cells, outcomes and regressors. The resulting 108 models have the following ranges across specifications, not confidence intervals.

Principal diagnosis codes	All-hospital offset ratio	Registered-birth offset ratio
Autism, F840/F841	1.030-2.196	1.027-2.048
Epilepsy/status epilepticus, G40/G41	0.964-1.166	0.987-1.095
Cerebral palsy, G80	0.778-0.890	0.800-0.871

The autism-coded associations remain positive with either denominator; they cannot be dismissed solely as an artifact of using hospital authorizations as the denominator. Differences from the preceding unrestricted-sample model ranges also reflect the restricted birth-cohort support, not merely the offset. The birth offset measures authorization intensity relative to registered births, not individual incidence or person-time: completed-age intervals cross calendar years, and

survival, migration, registration coverage and within-year birth timing remain relevant. A constant half-cohort offset leaves all 108 coefficients unchanged, but that invariance does not correct variable follow-up.

Separate reconstruction of the aggregate cells and a QR-orthogonalized Newton Poisson solver reproduce all 108 log coefficients within 0.000000002. Thirty-six deterministic known-mean recovery checks verify fitting implementation, not statistical power or interval coverage. Identification remains conditional on the restricted cohort counterfactual: an unrestricted birth-cohort shift can exactly offset any proposed vaccine coefficient while leaving fitted means unchanged. This algebraic equivalence does not establish the magnitude of actual confounding; it specifies the restriction needed to identify a vaccine effect. Dependence-aware inference, differential ascertainment and exposure measurement remain unresolved. These overlapping sensitivity models are not assigned an independent causal likelihood, and neither their positive nor their negative coefficients establish vaccine harm or protection.

Birth-registration coverage provides a further denominator sensitivity. The Ministry of Health's 2023 report, *Dados para Vigilancia*, Table 1 (page 16), reports national SINASC coverage of 95.2% in 2004 and 94.6% in 2007. Conditional on interpreting these published coverage estimates as registered divided by total births, the 2007-versus-2004 authorization-per-birth ratios are multiplied by 0.993697. The F840/F841 ratios consequently change from 9.770 to 9.708 at age five and from 0.274 to 0.272 at age ten. All 162 separated-cohort comparisons are retained, including 138 outcome-complete comparisons and 131 defined ratios. The published-coverage sensitivity leaves the contrasting age pattern intact. The source's 126 national and regional coverage cells for 2000-2020 were extracted and checked against PDF word positions; no values are extrapolated to earlier cohorts. The table provides no uncertainty intervals. This conditional adjustment is not an independent replication, does not resolve numerator ascertainment or cohort confounding, and cannot be applied as a single multiplier to the age-period regression coefficients. Source: Ministry of Health (2023).

The coverage-adjusted age-period regressions use common birth-cohort support of 2000-2017, without imputing coverage for 1996-1999. All eighteen specifications are fitted with each of three offsets for each diagnosis, producing 162 conditional models. Coverage-expanded births equal registered births divided by the published national coverage for that birth year. The resulting specification ranges are:

Principal diagnosis codes	All-hospital offset	Registered-birth offset	Coverage-expanded birth offset
Autism, F840/F841	0.993-2.088	0.991-2.048	0.975-2.046
Epilepsy/status epilepticus, G40/G41	0.946-1.171	0.961-1.070	0.953-1.072
Cerebral palsy, G80	0.768-0.939	0.783-0.885	0.775-0.874

These ranges are not confidence intervals. Some autism-coded specifications now lie slightly below one even before coverage adjustment; the difference from the 1996-2017 analysis therefore cannot be attributed entirely to registration coverage. Comparisons between columns isolate offset sensitivity on identical cells. A separate numerical solver reproduces all 162 log coefficients within 0.000000001, and all 54 deterministic mean-recovery and 162 constant-offset-rescaling checks pass. These verify implementation, not the absence of confounding or the reliability of conventional standard errors. The published coverage estimates are treated as fixed sensitivity inputs, without uncertainty propagation. No model in this extension contributes an independent causal likelihood.

An alternative Ministry coverage workbook, acquired on 8 September 2026 and labelled as finalized through 2024, provides a separate source-version sensitivity. On the same 2000-2017 cohort cells, coverage-adjusted specification ranges are 0.992-2.063 for F840/F841, 0.952-1.079 for G40/G41 and 0.774-0.886 for G80. Relative to the 2023-report adjustment, the 54 paired ratios change by -0.276% to +1.759%. All 162 refits are independently reproduced; 108 hospital- and registered-birth-offset coefficients remain unchanged and are duplicate checks, not additional evidence. The current source uses projection inputs through 2014 and capture-recapture inputs from 2015 onward. Its coverage indices can exceed 100% at subnational levels and are not uniformly registration probabilities. Both source versions are retained, without selecting one according to its coefficient. These comparisons do not propagate coverage uncertainty or resolve age-period-cohort identification. Source: Ministry of Health, coverage and correction-factor workbooks.

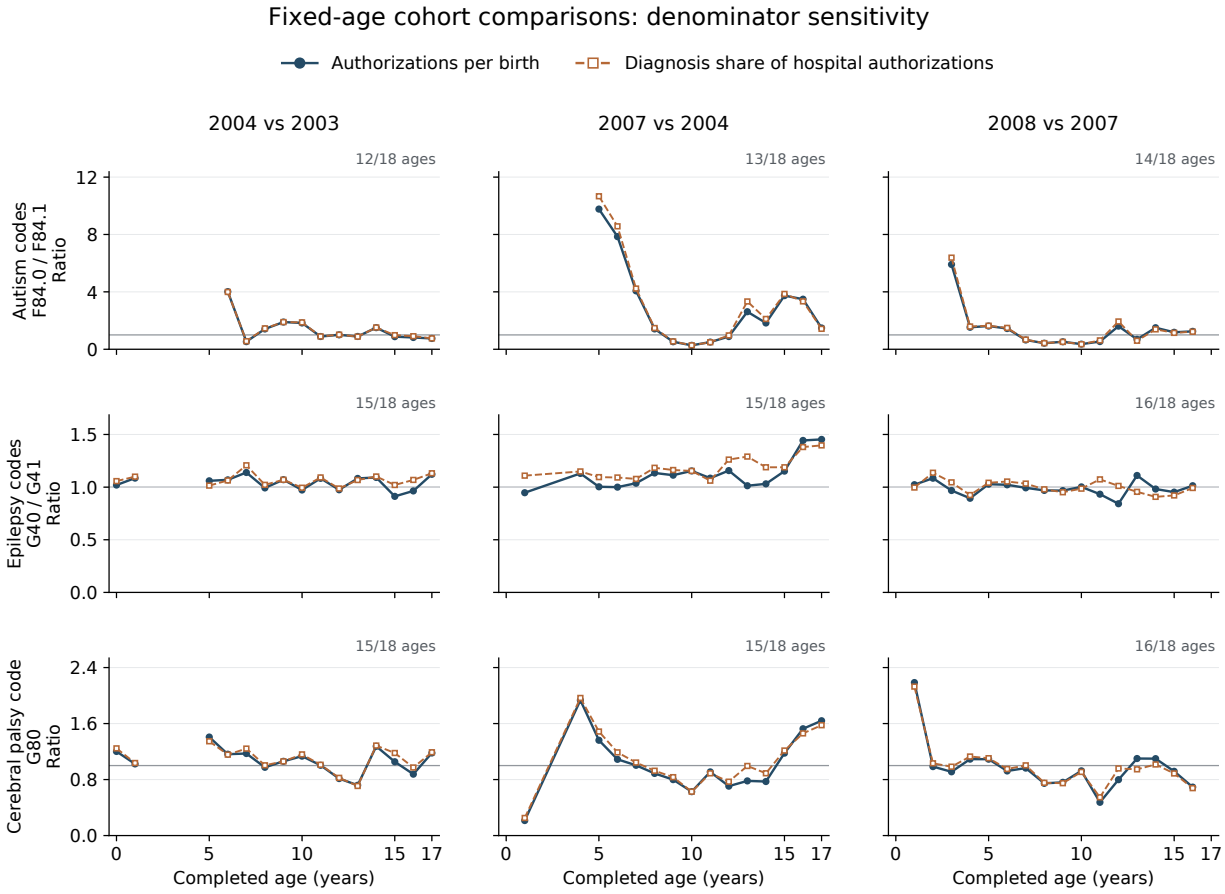
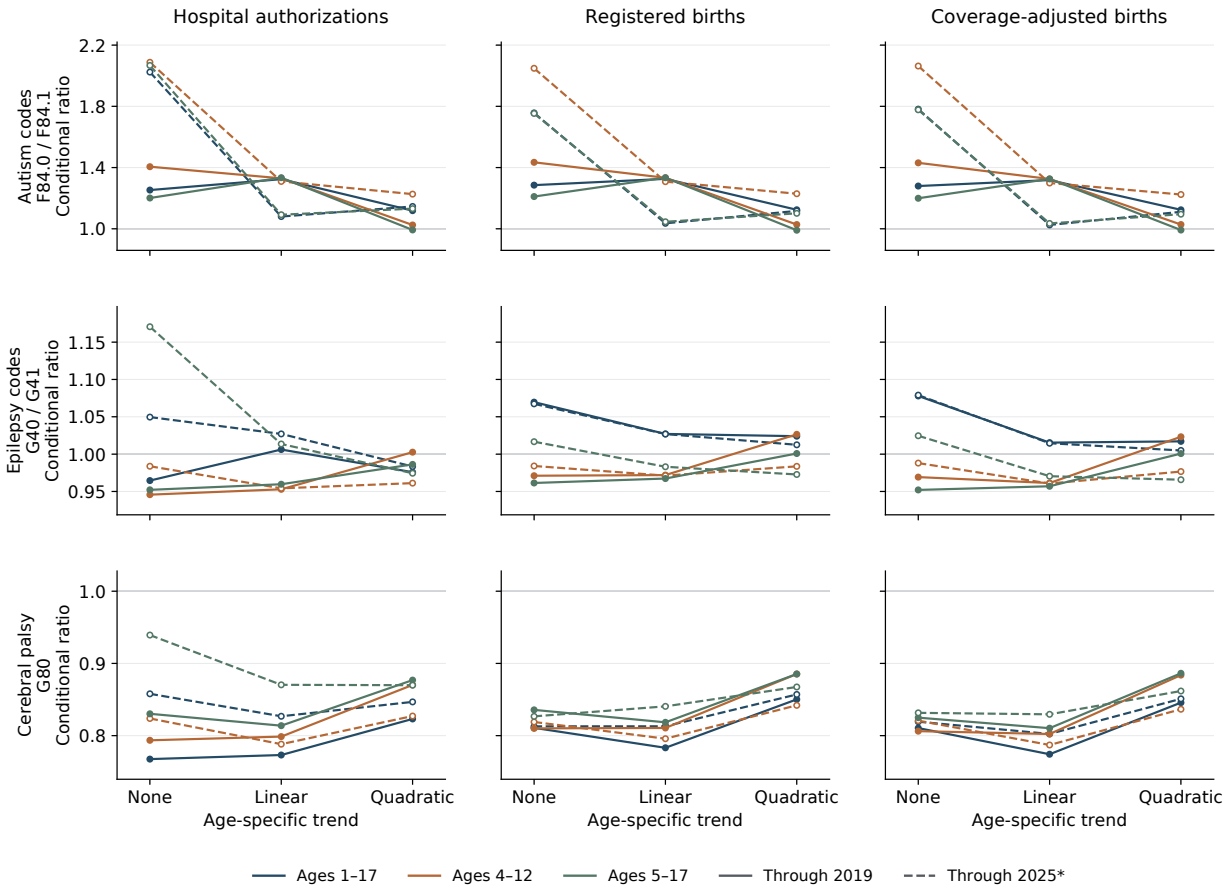


Figure S52. Fixed-age cohort comparisons under birth and hospital denominators. All 131 defined ratios from 162 attempted contrasts are shown; gaps indicate incomplete windows or zero baseline counts. The outcomes are principal-diagnosis hospital authorizations, not unique diagnosed individuals. Both series use the same records and are not independent replications. Row scales differ. These descriptive comparisons do not identify vaccine effects.

Paired specification comparisons separate denominator changes from trend and sample changes. For F840/F841, adding age-specific linear trends changes the fitted ratio by -47.17% to +11.05% across eighteen matched pairs; adding quadratic terms changes it by -25.71% to +8.36%. Extending follow-up from 2019 to 2025, excluding 2020-2022, changes it by -22.22% to +72.09% across twenty-seven pairs. In comparison, replacing registered births with current coverage-adjusted births changes it by -1.03% to +1.39%, and changing coverage-report vintage changes it by +0.03% to +1.73%. These percentages describe changes in coefficients, not percentage effects on autism and not confidence intervals. Follow-up comparisons also change the observations. The complete paired accounting includes age windows and cohort-support changes, does not select a preferred model by proximity to one, and does not count repeated specifications as independent evidence. Thus, denominator refinement does not resolve the substantially greater sensitivity to cohort-trend restrictions and follow-up support.



All 162 matched fits; birth cohorts 2000–2017. Shared data, not independent studies. Row scales differ.
*2020–2022 excluded. Conditional authorization ratios—not individual risks; no inferential intervals.

Figure S53. Trend sensitivity on matched cells. All 162 conditional fits use birth cohorts 2000-2017, three age windows, two follow-up endpoints, three age-specific trend degrees and three denominators. Lines connect specifications, not time observations or independent studies. The longer window excludes 2020-2022. Row scales differ. These authorization ratios do not identify individual vaccine risks, and no inferential intervals are shown.

S45.5 Pennsylvania age-specific service-share analysis

PennData Table 10A reports disability counts by single age and Intermediate Unit (IU), with separately printed all-disability age totals. We use those printed totals as denominators: summing only visible disability categories would omit masked cells and distort the service share. This analysis covers 1,547 IU-year-age observations in 28 IUs, ages 5-15, during 2015-2024. Printed totals are available for every retained observation. Autism counts agree with direct re-extraction throughout. Independent extraction verifies all 1,547 denominator cells: text-order extraction covers 892 cells in four modern report vintages, and a separate PDF engine verifies the remaining 655 cells across 161 older-layout pages.

Poisson models relate autism counts to kindergarten MMR coverage assigned to the corresponding birth cohort, with the printed disability count as an offset. All specifications include IU and age fixed effects. The two time adjustments below impose different restrictions and are presented together. Ratios describe autism's share among disability-service recipients per 10 percentage-point higher coverage, not population autism risk.

Geographic support and time adjustment	Service-share ratio	95 percent interval
All 28 IUs, birth-cohort fixed effects	0.868	0.725-1.040
All 28 IUs, calendar-year fixed effects	0.898	0.841-0.959
Unanimous matched-LEA assignments, cohort effects	0.931	0.623-1.393
Unanimous matched-LEA assignments, year effects	0.959	0.881-1.043

The table uses enrollment-weighted coverage reconstructed from vaccinated and enrolled counts for all 67 counties in each of 12 school years, 2014-2025 (804 county-years). County coverage is assigned to the IU serving the plurality of matched LEAs. Weights are county kindergarten enrollments, not the fractions of pupils residing in each IU. LEA membership shares are not pupil shares. Requiring unanimous matched-LEA assignments for every county assigned to an IU retains 877 observations in 16 IUs; both time-adjusted intervals then include one. This restriction excludes entire IUs rather than dropping exposure counties while retaining the same outcome population. It changes the sample and does not establish the crosswalk's historical accuracy. Equal county weighting gives cohort-adjusted ratios of 0.919 (0.750-1.126) in all IUs and 1.017 (0.705-1.468) under unanimous matching. All these comparisons use the same selected outcome cells.

Intervals use IU-cluster standard errors with a t critical value and one fewer degree of freedom than the retained number of IUs. This accommodates within-IU dependence but is not a guarantee against common cross-IU shocks or few-cluster distortion. These are correlated sensitivity estimates, not additional independent studies. Negative associations do not establish protection: coverage assignment, selection into services, and age-specific diagnostic trends remain potential explanations.

This age-specific analysis is distinct from the 67-county PennData series used in the local-MMR sensitivity synthesis. The two Pennsylvania series must not be substituted for one another or treated as independent population samples. The age-comparison analysis module includes scripts for printed-total extraction, matched-sample refitting, and independent verification. Source-PDF hashes and page references accompany the reconstructed observations.

S45.6 Reporting-vintage and missing-outcome checks

The exposure reconstruction permits the following reporting-year exclusions, using enrollment weighting and birth-cohort fixed effects in all 28 IUs. Exclusions are specified for source-scope and service-disruption concerns, not selected by their estimated signs. Each changes the observation window rather than supplying an independent replication.

Excluded reporting years	IU-year-age cells	Service-share ratio	95 percent interval
None	1,547	0.868	0.725-1.040
2019	1,396	0.869	0.728-1.037
2020-2021	1,192	0.872	0.734-1.035
2019-2021	1,041	0.873	0.736-1.035

The analysis module retains all 32 combinations of four reporting windows, two exposure-weighting rules, two geographic-support rules and two time adjustments. Eight reproduce the full-window fits; 24 are exclusion sensitivities. Removing pandemic reporting years does not isolate a pandemic causal effect, and none of these specifications supplies an exogenous vaccination first stage.

Source inspection distinguishes 33 additional masked autism cells from 166 candidate records whose ages are absent from the printed source header. The latter are not censored observations, zeros, or additional sample size. Independent extraction verifies the 33 masks and their printed disability denominators. No imputed outcomes are included in the estimates above.

The 2019 autism row totals have a scope inconsistency: among 26 included IU pages, all 24 with visible age-five counts have a Table 1 count exactly equal to the Table 10A total plus age five, despite Table 10A's 5-21 heading. The remaining two age-five counts are masked. Accordingly, no 2019 row total is used to constrain masked cells, even when subtracting visible entries would yield a nonnegative residual. Provisional count upper constraints for the 33 masks range from 15 to 167, not a uniform ten; these are neither confidence intervals nor causal-effect bounds. Any use of other vintages' row totals requires compatible age definitions and joint reconciliation across all printed ages. Suppressed-count estimation is not part of these reported fits.

S45.7 Chile census disability and exploratory cohort profiles

Chile's 2024 public census person file contains 18,480,432 records with age, sex, commune and six functional-difficulty responses. These are functional outcomes, not autism diagnoses. Applying the sequential disability rule in the INE microdata manual (page 94) reproduces the supplied classification for every record. The resulting counts are 1,950,388 people with disability, 15,561,356 without disability, 97,995 nonresponses and 870,693 children aged 0-4 for whom the measure does not apply. Missing responses are not coded as absence of disability.

Descriptive calculation	Disability percentage
Among classified respondents	11.137600%
Complete responses to all six domains	11.135434%
All nonrespondents assigned without or with disability	11.075621% to 11.632103%
Upper bound also allowing severe difficulty in partially observed negative records	11.722610%

The final row includes 15,938 partial-response records classified without disability by the published rule. These ranges concern missing answers, not sampling uncertainty, vaccine effects or errors in observed answers. The analysis retains 204 five-year-age-by-sex-by-domain summaries. Different domains and age groups reuse the same census collection and do not constitute independent studies.

The single-year-age screen includes 17,192,683 people aged 5-84 in 295 communes. In the remaining 51 communes, 151,199 people have age suppression code -66 and require the separate five-year age-band field. Ages 85 and above are top-coded. Completed-year age does not identify exact birth date. Geographic-layer population totals also differ from the person file; the geographic and person-level denominators are not substituted for one another.

An exploratory Hib cohort-profile screen considers the July 1996 infant-program introduction. MINSAL specifies doses at two, four and six months; Chile's periodic report to the Committee on the Rights of the Child specifies infants turning two months from July onward. Prior selective Hib vaccination in Santiago and a 1996 measles campaign complicate attribution. Neither an unexposed comparison cohort nor a measured uptake first stage is established.

Respondent-weighted linear and quadratic age profiles, with common or separate slopes, are fitted for both sexes pooled and separately. Omitting ages 27-28 and retaining six, eight or ten annual cells per side gives 36 specifications. The modeled younger-minus-older step at age 27.5 ranges from -0.248 to +0.287 percentage points. Expanding the omitted interval to four or six age cells gives 108 target specifications with a range of -0.923 to +0.924 points. Ten shifted age cutoffs produce 360 further diagnostics, ranging from -1.444 to +2.665 points. These shifted cutoffs are not randomized placebo assignments, and their distribution does not supply a randomization p-value or false-positive rate.

Allowing nonrespondents' outcomes to vary adversarially by age in all-eligible-denominator fits produces coefficient ranges containing zero in every target specification. Direct fits verify all 936 bound endpoints across target and shifted specifications. The calculation does not address confounding, partial negative responses or reporting error. Extrapolation across omitted ages and the inability of one cross-section to separate unrestricted age and cohort effects preclude a causal interpretation. The screen contributes no causal pooling weight, individual vaccine-risk bound, Bayes factor or claim of excluding small effects.

Sources: INE download catalogue, INE microdata manual, MINSAL Hib history, Chile periodic report, paragraphs 635-636, and prior selective Hib vaccination study. Public-input reacquisition and isolated execution reproduce the five person-analysis outputs byte-for-byte; this computational check does not establish identification.

S45.8 Italian district exposure measurement and cohort checkpoints

The Emilia-Romagna coverage reports supply district objective tables for 2007–2008 and 2009–2010. Each two-year extract contains 928 printed percentages and 768 distinct cells after reconciling repeated regional margins. Together they describe 36 separately named districts, eleven health authorities and eight reported checkpoint measures per edition. Imola and Forli appear as authority totals without separately named district rows. Districts, authority totals and the regional total are nested, not independent observations. All 22 table pages were visually checked; ordered labels and values agree between two PDF readers. The original 46-page front-matter candidate inventory has extracted records for every listed page. That bounded accounting does not establish exhaustive Italian or worldwide acquisition.

The school-age measles checkpoint changes from age six in the 2008 report to age seven in the 2010 report. Concatenating those percentages as a constant-age series would mix cohort progression with calendar changes. The mandatory-vaccination measure records joint completion of polio, diphtheria, tetanus and hepatitis B vaccination, not a sum of marginal percentages, simultaneous administration or cumulative dose count. The abbreviated school-age one-dose heading denotes at least one dose rather than exactly one dose; vaccination coverage must also be distinguished from immunity after prior infection.

Age six in 2008 and age seven in 2009 identify the same nominal birth year, 2002, but not necessarily the same resident children. Across the 36 districts, the two-dose percentage increases by more than 0.1 points in 26, changes within 0.1 points in four and decreases by more than 0.1 points in six. The unweighted median change is +1.3 percentage points, with a range of -5.4 to +41.3. These are exposure-checkpoint descriptions, not vaccine effects. A decline beyond the conservative rounding envelope can reflect migration, denominator composition, recording or corrections; it does not establish vaccination reversal or an erroneous source. The report describes catch-up activity, so later-age coverage is not interchangeable with receipt at the scheduled age.

Detailed-table crosswalks compare 192 age- and geography-matched measles cells in each edition. All 192 printed percentages agree in the 2008 report; 190 agree in the 2010 report. The two exceptions concern Correggio. At 24 months, the summary gives 96.2%, matching immunity including a prior-infection case, whereas vaccinated children number 640 of 666, or 96.1% as printed. At age thirteen, the summary gives 99.0%, whereas the detailed table gives 504 vaccinated children among 507 residents, printed as 99.4%. Both versions are retained with source-page flags rather than silently harmonized. Count-derived percentages agree with the detailed table's rounding. These matches are repeated measurements from the same report and populations, not independent replication.

The inspected autism-service outcome tables provide regional age groups and broader authority-level outcomes, not a district-by-age outcome panel matched to these exposure checkpoints. The district detail therefore does not itself identify another vaccine effect or increase the effective sample size of the existing authority-level models. Geographic boundaries, diagnosis production, follow-up and cohort eligibility must be reconciled before any such join. No causal likelihood, individual RR/OR, Bayes factor or power gain is assigned to these exposure checks.

Sources: Emilia-Romagna 2008 coverage report, summary pages 22–32 and detailed appendix and 2010 coverage report, summary pages 23–33 and detailed appendix. The source catalog and evidence register link these reports to the extraction scripts, source hashes and table crosswalks.

S45.9 Source-to-panel reproduction

The Brazilian hospital-input reconstruction processes all 8,099 held state-month archives used for the 2001–2025 residence panels. The 25 annual panels reproduce every reference cell, totaling 22,336,849 aggregate panel rows. The underlying files contain 300,608,590 active administrative authorization records, not unique patients or independent observations. The missing Amapa October 2007 archive is not imputed as zero; design-specific exclusions and matched-month sensitivities remain necessary. Terminal verification checks file identities and source hashes, annual inventories, status partitions, retained-row accounting, generated-file hashes and unchanged recorded code and manifest inputs. The compiled decompressor was separately verified and its final hash checked; its binary was not continuously inventoried throughout the full run.

The standalone Italian exposure module downloads four official reports and reproduces seven extraction outputs exactly, including both district objective appendices. It checks the 52 pinned dependencies and completes fifteen extraction, reconciliation and sensitivity steps. Separate tests reject changed code, a mismatched runtime pin, corrupted cached data and missing, stale or failed validation records. The detailed-table crosswalk is a separate working-copy dependency, not part of that source-free module. The public module index records the exact archives and bounded execution receipts. Neither exact numerical reproduction nor the volume of administrative records establishes causal identification, independent population replication or an individual-risk bound; full-manuscript reproduction has a wider dependency scope than these modules.

S46. Power on ecological and individual scales

The PCV7 coefficient-to-risk mapping failed validation (S52); its individual-equivalent power and posterior values are not reported here. The RotaShield illustrations remain conditional rather than identified risk probabilities. Section S52 shows that the spline coefficient is not the exposure contrast assumed by the mixture equation. Its model-native power must therefore be distinguished from exposure-aligned power against an individual effect.

S46.1 Exposure contrast and power

A detectable change in a policy contrast is not the same as a detectable vaccinated-versus-unvaccinated risk difference. We illustrate the distinction using the fixed-share homogeneous-mixture forward maps from the individual-risk calibration. For individual risk ratio R and exposure shares pH and pL, the mapped ecological ratio is Q = [1 + pH(R - 1)] / [1 + pL(R - 1)]. This mapping requires a common baseline risk, homogeneous effects, valid exposure shares, no interference and no residual ecological confounding; it does not establish these assumptions.

For transparency, the following power approximation uses only the rounded native 95% intervals: RotaShield, 0.945-1.051. The log standard error is the difference between the logged interval endpoints divided by twice the 97.5th standard Normal percentile. Two-sided rejection probability at alpha 0.05 is evaluated under each mapped alternative, against a zero ecological coefficient rather than against the observed point estimate.

Design	Hypothetical individual RR	Mapped ecological RR	Approximate power
RotaShield uptake	1.05	1.0050	5.4%
RotaShield uptake	1.10	1.0100	6.6%
RotaShield uptake	1.20	1.0200	11.3%

The corresponding approximate 80%-power threshold is ecological RR 1.079 for RotaShield, not an individual-risk threshold. These calculations do not additionally propagate exposure-share uncertainty, allow residual bias or heterogeneous effects, calibrate finite-cluster test size, or adjust for multiplicity. They are an illustration from rounded inputs, not an empirical power simulation, a rerun of the original power archive, or additional independent evidence. They show why narrow ecological intervals alone cannot exclude small individual effects.

The forward-map report's seven archived input hashes match the current input files. This is a provenance check, not numerical reproduction of that report or validation of its causal assumptions. The calculation and machine-readable results are provided in `qa/completeness_20260907/age_control_designs/calibrate_power_scale_illustration.py` and `power_scale_illustration.json`.

S46.2 Posterior integration and tail sensitivity

Exposure Monte Carlo stability is assessed with three additional seeds and 999 draws per design per seed, totaling 5,994 model refits. Nested prefixes of 249, 499 and 999 draws produce 72 posterior summaries across the two designs and four priors. At 999 draws, the largest absolute difference from the original probability above individual RR 1.10 is 0.41 percentage points. The PCV7 rows in this archived numerical audit use an invalid individual mapping and have no inferential role. Under the neutral prior, RotaShield medians range from 0.9493 to 0.9545 and probabilities from 28.84% to 29.55%. All draws succeed. These seed ranges describe numerical sensitivity, not confidence intervals, additional participants, or proof of convergence. They remain conditional on the independent Beta/Kish exposure approximation and the specified outcome models and bias allowances.

A separate same-seed rerun starts from the seven source-input tables and regenerates all 498 exposure-draw model fits, 16 fixed-share and joint posterior rows, eight forward/inverse mappings, and 11 validation rows (the PCV7 rows among them are withdrawn). Maximum absolute numerical differences from the archived draws and posterior tables are below 0.000000000000003; the mapping and validation tables agree exactly. The rerun records input and analysis-module hashes, preserves the archived files, and writes to an isolated audit directory. This establishes computational reproduction of this calibration module, not independent confirmation of its estimator or sufficiency of its exposure-draw count.

An independently implemented likelihood reproduces all eight joint-calibration posterior medians and probabilities above individual RR 1.10 (the four PCV7 rows are withdrawn) within 0.000000001, conditional on the 498 saved exposure draws. Adaptive integration over the entire real log-risk line removes the original grid's bounds of -3 and 3 and its discretization. The largest posterior mass outside those bounds is approximately 0.0083%, for RotaShield under the Student-t prior. Tail calculations also examine log-risk bounds of 6 and 12. Negligible tails at machine precision are numerical approximations, not proof of bounded risk.

Design	Prior	Individual RR median	95% credible interval	P(individual RR > 1.10)
RotaShield	Tight skeptical	0.990	0.715-1.347	25.6%
RotaShield	Neutral	0.952	0.520-1.528	29.2%
RotaShield	Harm-favoring	1.030	0.581-1.614	39.7%
RotaShield	Student-t, 3 degrees of freedom	0.949	0.472-1.523	28.5%

The Normal priors have log-risk centers and standard deviations of (0, log(1.20)), (0, log(1.50)), and (log(1.20), log(1.50)); the Student-t prior has location zero and scale log(1.50). These continuous priors assign zero probability to an exact point null. Any mass at the zero grid point is a numerical cell, not a posterior probability that the effect is exactly zero. Point-null probabilities require a separately specified mixture prior. The table reports conditional ecological-to-individual extrapolations, not established causal harms or independent evidence. This check verifies integration conditional on archived draws, not the source-model fits, adequacy of 249 draws per design, or causal identification. Broad intervals and prior sensitivity remain after numerical truncation is removed.

S47. Childhood diabetes: EURODIAB introduction-date and age-gradient comparisons

Patterson et al. (2019; doi:10.1007/s00125-018-4763-3), Table 1 and electronic supplementary Table 2, provide 124 observed centre-period standardized incidence rates and 744 age-by-sex rates from 26 centres in 22 European countries. Six centre-periods and their 36 age/sex cells are missing, not zero. Equal-weight means of the six strata reproduce every published standardized rate within rounding tolerance. Actual partial-period observation years and the 1999 catchment expansions in Denmark and North Rhine-Westphalia are retained. These are repeated measurements of the same populations, not independent policy experiments.

Introduction-date models use the 124 centre-period standardized rates directly. Each regresses log incidence on centre and period fixed effects plus one exposure term, with country-clustered CR1 covariance and t(G-1) critical values. Midpoint coding marks a centre-period as exposed when its midpoint year is at or after national introduction plus the lag. Calendar-fraction coding uses the share of the period's observed calendar years at or after that date. Hib is fitted at seven- and ten-year lags and PCV at a seven-year lag. Each fit uses all available centres and is repeated excluding the two centres whose catchments changed in 1999. Neither coding measures uptake or exposed birth-cohort person-time. The twelve specifications are correlated variants, not replications. No pretrend, placebo-date or heterogeneity-robust estimator was fitted for this family.

Table S47-A. Introduction-date specifications.

Antigen	Lag (years)	Sample	Timing	RR (95% CI)	p	Centres	Countries (df)	Obs.	Exposure-coded obs. (centres)	Sum of exposure fractions
Hib	7	all available	midpoint	0.892 (0.785 –1.012)	0.075	26	22 (21)	124	58 (23)	58.0
Hib	7	all available	calendar fraction	0.833 (0.724 –0.959)	0.013	26	22 (21)	124	66 (24)	58.2
Hib	7	excl. changed catchments	midpoint	0.888 (0.781 –1.010)	0.068	24	21 (20)	114	52 (21)	52.0
Hib	7	excl. changed catchments	calendar fraction	0.826 (0.717 –0.953)	0.011	24	21 (20)	114	59 (22)	52.0

Antigen	Lag (years)	Sample	Timing	RR (95% CI)	p	Centres	Countries (df)	Obs.	Exposure-coded obs. (centres)	Sum of exposure fractions
Hib	10	all available	midpoint	0.864 (0.742 –1.007)	0.060	26	22 (21)	124	43 (20)	43.0
Hib	10	all available	calendar fraction	0.814 (0.680 –0.974)	0.026	26	22 (21)	124	52 (21)	44.2
Hib	10	excl. changed catchments	midpoint	0.855 (0.731 –1.001)	0.051	24	21 (20)	114	38 (18)	38.0
Hib	10	excl. changed catchments	calendar fraction	0.806 (0.672 –0.967)	0.023	24	21 (20)	114	46 (19)	39.2
PCV	7	all available	midpoint	0.900 (0.660 –1.225)	0.483	25	21 (20)	119	3 (3)	3.0
PCV	7	all available	calendar fraction	0.843 (0.574 –1.239)	0.366	25	21 (20)	119	14 (14)	5.4
PCV	7	excl. changed catchments	midpoint	0.897 (0.655 –1.228)	0.478	23	20 (19)	109	3 (3)	3.0
PCV	7	excl. changed catchments	calendar fraction	0.834 (0.560 –1.242)	0.351	23	20 (19)	109	13 (13)	5.2

Entries are ratios per unit of the coded exposure term, not vaccinated-versus-unvaccinated RRs. All four Hib calendar-fraction intervals lie below one (estimates 0.806–0.833). The four Hib midpoint estimates, 0.855–0.892, have upper limits of 1.001–1.012. The four PCV estimates, 0.834–0.900, have intervals spanning 0.560–1.242. PCV fits use 25 centres in 21 countries because no PCV date is coded for Czechia. At the seven-year lag, only 3 PCV midpoint observations (3 centres) are coded exposed. Calendar-fraction coding gives 13–14 partially exposed observations whose fractions sum to 5.2–5.4 observation-equivalents. No leave-one-country-out analysis exists for this 26-centre introduction-date family; the 86 country-deletion fits after Table S47-B concern the age-gradient models. In the earlier 21-centre sample (20 centres in 17 countries for PCV), the PCV seven-year-lag midpoint estimate of 0.884 (0.558–1.401) ranged across 17 single-country deletions from 0.687 (0.615–0.768) without Austria to 1.137 (1.017–1.271) without Italy. The inverse Hib intervals are not evidence of protection. All twelve specifications use the inherited introduction dates in Table S47-C, including dates the partial source review found incomplete.

Table S47-C. Partial review of inherited introduction dates.

Country / centre, antigen	Inherited year (used in Table S47-A)	Reviewed programme year	Review finding	Historical uptake verified
Poland, Hib	2004	2007	2004 applied to young-child institutional settings; the universal programme began in 2007. Earlier vaccination must not be coded as zero, but 2004 is not a universal introduction.	no
Sweden, Hib	1992	1993	The historical programme table specifies 1993, with doses at 3, 5 and 12 months; a separate agency summary gives 1992. The discrepancy is retained as a timing sensitivity.	no

Country / centre, antigen	Inherited year (used in Table S47-A)	Reviewed programme year	Review finding	Historical uptake verified
Sweden (Stockholm County), PCV	2006	2007 (births from 1 July 2007)	Stockholm implemented PCV7 for infants born from July 2007, with doses at 3, 5 and 12 months. The regional date, not the national 2009 programme, matches this centre.	no
Finland, PCV	2009	2010 (1 September 2010)	Routine PCV10 began in September 2010. Earlier trial exposure needs separate treatment, not erasure or universal-exposure coding.	no
United Kingdom, Hib	1992	1992 (October)	Infant doses at 2, 3 and 4 months were accompanied by a one-dose catch-up for children under four, so eligibility is not restricted to births after 1992. A 2003 catch-up campaign is a further exposure event.	no

The review is partial (`partial_source_review_not_complete_exposure_validation`). The remaining country-product pairs, exact historical eligibility and uptake are unreviewed. The age-gradient models below use both the inherited dates and a partially reviewed scenario built from these findings.

The age-gradient outcome is log incidence at ages 0-4 or 5-9 minus log incidence at ages 10-14 within centre, sex and period. The regressor is the difference in modelled programme eligibility between those age groups. Monthly age/calendar grids assume uniform person-time within cells and a provisional routine administration age of six months. Older children are allowed to be eligible. The exposure scenarios compare inherited introduction dates with a partially source-reviewed set that distinguishes Polish Hib partial and universal rollout, Stockholm PCV eligibility, Finnish PCV routine introduction and UK Hib catch-up; Swedish Hib timing disagreement is retained across scenarios. These models do not measure actual receipt, fully resolve historical catch-up, or establish a first stage.

The 32 exploratory specifications cross two antigens, two exposure scenarios, two age comparisons, zero/three-year additional latency and inclusion/exclusion of centre-by-sex linear trends. All include centre-by-sex and period-by-sex fixed effects. Country-clustered CR1 covariance and t critical values with G-1 degrees of freedom provide approximate inference. Six nominal associations have p below .05; none survives Holm adjustment across this specification family. Some older-age latency specifications are numerically identical, and the family is not a set of independent replications.

Exposure	Ages compared	Without centre-specific linear trends	With centre-specific linear trends
Hib	0-4 versus 10-14	1.124 (0.982-1.286)	0.951 (0.791-1.144)
PCV	0-4 versus 10-14	0.721 (0.540-0.963)	1.134 (0.751-1.712)

The table shows ratios of age-specific rate ratios per unit modelled eligibility contrast, using the partially reviewed scenario and zero additional latency. They are not individual vaccinated-versus-unvaccinated RRs. Sign and precision changes under trend adjustment preclude either a protective interpretation or a precise null conclusion. Stratum-specific counts and person-years are not supplied by the rate table and are not reconstructed as pseudo-counts. All these estimates receive zero causal-synthesis weight; cohort composition, uptake, pre-period support, alternative counterfactual trends and ascertainment remain identification requirements. Reproduction scripts are `processed/eurodiab_complete_table_audit.py`, `processed/eurodiab_age_sex_extract.py` and `processed/eurodiab_age_contrast.py`; full results, source hashes and extraction checks are archived in `qa/claude_corrections_20260919/`.

Table S47-B. All 32 age-gradient specifications, as no-trend versus centre-by-sex-trend pairs.

Antigen	Dates	Added latency (years)	Ages compared	Without trends	With trends	Crosses one?
Hib	inherited	0	0–4 versus 10–14	1.123 (0.990–1.274)	0.953 (0.796–1.140)	yes
Hib	inherited	0	5–9 versus 10–14	0.945 (0.809–1.104)	0.917 (0.747–1.126)	no
Hib	inherited	3	0–4 versus 10–14	1.274 (1.013–1.602)	0.934 (0.712–1.227)	yes
Hib	inherited	3	5–9 versus 10–14	0.945 (0.809–1.104)*	0.917 (0.747–1.126)*	no
Hib	partially reviewed	0	0–4 versus 10–14	1.124 (0.982–1.286)	0.951 (0.791–1.144)	yes
Hib	partially reviewed	0	5–9 versus 10–14	0.952 (0.800–1.134)	0.940 (0.743–1.189)	no
Hib	partially reviewed	3	0–4 versus 10–14	1.276 (1.009–1.614)	0.944 (0.712–1.251)	yes
Hib	partially reviewed	3	5–9 versus 10–14	0.949 (0.796–1.131)	0.937 (0.739–1.188)	no
PCV	inherited	0	0–4 versus 10–14	0.693 (0.516–0.929)	1.128 (0.733–1.735)	yes
PCV	inherited	0	5–9 versus 10–14	1.044 (0.701–1.554)	1.356 (0.728–2.528)	no
PCV	inherited	3	0–4 versus 10–14	0.378 (0.161–0.889)	1.419 (0.339–5.942)	yes
PCV	inherited	3	5–9 versus 10–14	1.044 (0.701–1.554)*	1.356 (0.728–2.528)*	no
PCV	partially reviewed	0	0–4 versus 10–14	0.721 (0.540–0.963)	1.134 (0.751–1.712)	yes
PCV	partially reviewed	0	5–9 versus 10–14	1.092 (0.722–1.650)	1.410 (0.742–2.681)	no
PCV	partially reviewed	3	0–4 versus 10–14	0.423 (0.181–0.989)	1.475 (0.344–6.322)	yes
PCV	partially reviewed	3	5–9 versus 10–14	1.092 (0.722–1.650)*	1.410 (0.742–2.681)*	no

* Numerically identical to the corresponding zero-latency fit.

Entries are ratios of age-specific rate ratios per unit of modelled eligibility contrast (95% intervals), not vaccinated-versus-unvaccinated RRs. The six starred lag-3 fits at ages 5–9 duplicate their lag-0 counterparts, so the 32 specifications contain 26 distinct results. In all eight ages 0–4 pairs, adding centre-specific trends moves the estimate to the other side of one; this happens in none of the eight ages 5–9 pairs. Six specifications have $p < .05$: two harmful-direction Hib estimates, 1.274 (1.013–1.602) and 1.276 (1.009–1.614), and four inverse PCV estimates, 0.693, 0.378, 0.721 and 0.423. None survives Holm adjustment across the 32 specifications; the smallest Holm-adjusted p is 0.538. The two-row table above repeats the partially reviewed, zero-latency 0–4 pairs.

Leaving countries out one at a time adds 86 influence fits for the partially reviewed, zero-lag, ages 0–4 versus 10–14 contrasts. Hib point estimates range from 1.082 to 1.148 without centre-specific trends and 0.928 to 0.986 with them; PCV ranges are 0.687–0.777 and 1.030–1.216, respectively. These are ranges of correlated point estimates, not confidence intervals. Two no-trend Hib deletions yield nominally positive intervals (omitting Luxembourg or Switzerland), and 16 no-trend PCV deletions yield nominally inverse intervals; none of the trend-adjusted deletion intervals excludes one. Each specification retains its point-estimate direction after every deletion, so a single country does not explain the disagreement between trend specifications. This does not identify the correct counterfactual or repair exposure measurement. All deletion results retain zero causal-synthesis weight. The script is `processed/eurodiab_age_country_influence.py`.

S48. Published-rate chronic-outcome sensitivities

YRBS is school-based: it excludes homeschooled and out-of-school youths, and participating jurisdictions differ in covered school types. The national and site samples can share respondents. DNPAO and DASH publications derived from YRBS are therefore not independently sampled replications, and their sample sizes must not be added as unique people. Attendance-related selection is especially relevant to disability outcomes. Survey weights address sampling and nonresponse; the aggregate regressions below do not recover the full survey design (CDC, 2023 YRBS methods).

The youth obesity and asthma series retain state/year/grade cells; the WIC series retains state/year/age-band cells. Suppressed observations remain missing. Weighted prevalence multiplied by sample size is not treated as an observed case count, and repeated-wave sample totals are not treated as unique participants. Asthma denotes self-reported lifetime diagnosis by a health professional, not incident disease. WIC obesity pertains to participating toddlers, not a representative sample of all children.

Log prevalence is regressed on state rollout intensity interacted with the proxy-cohort exposure window, with state, proxy-cohort, survey-year and grade/age-band main effects. SVD handles redundant nuisance columns and exposure rank is checked. Equal-cell, sample-size and inverse squared log-CI-width weighting are compared, with proxy birth-cohort shifts of minus one, zero and plus one year. The CI-based weights approximate precision but do not recover the survey design. State-clustered CR1 covariance and t(G-1) critical values provide conditional sampling intervals, not protection from confounding or cohort misclassification.

Outcome / exposure contrast	Equal-cell ratio (95% interval)	Sample-size weighted	Published-CI weighted
Youth obesity / PCV7 step	0.970 (0.954–0.985)	0.975 (0.961–0.989)	0.969 (0.954–0.985)
Youth obesity / RotaShield 1998 pulse	1.013 (0.977–1.051)	1.016 (0.993–1.039)	1.014 (0.992–1.035)
Ever-diagnosed asthma / PCV7 step	0.988 (0.959–1.019)	0.979 (0.951–1.008)	0.978 (0.951–1.005)
WIC toddler obesity / rotavirus reintroduction	1.014 (0.996–1.033)	1.015 (0.991–1.040)	1.013 (0.992–1.035)

The table shows zero-shift estimates per standard deviation of state rollout intensity, not individual vaccine-exposure RRs. Youth obesity uses 48 state clusters. Asthma uses 947 cells in 46 states; WIC uses 1,071 cells in 51 state/DC units. The 18 youth-obesity and 18 asthma/WIC specifications are correlated sensitivity fits, not 36 independent studies.

Estimator-matched youth-obesity diagnostics use pre-event cohorts, an intensity-by-linear-cohort slope and a 1993 placebo step or pulse. Across the nine PCV7 specifications, estimates range from 0.962 to 0.976, every interval below one. All nine pretrend p values are below .006, with intensity-by-cohort slopes across the pre-PCV7 cohorts born 1985–1999 of -0.00262 to -0.00446 log points per standard deviation per cohort-year, and all nine placebo p values are below .015. At zero shift with equal-cell weighting, the fictitious 1993 step is 0.967 (0.953–0.982), about as large as the main estimate. Thus the PCV7 inverse association is not interpreted as prevention. RotaShield diagnostics do not reject at .05, but shifting the proxy cohort by minus one year makes all three weighting rules nominally inverse (equal-cell 0.975, 0.951–0.998; p = .002–.041); with surveys ending in 2019, one of the three remains so. Asthma's unshifted pretrend p values range .678–.883 and placebo p values .325–.476; non-rejection does not establish diagnostic power or parallel counterfactual trends.

WIC has only three possible pre-event proxy cohorts in this window. Its diagnostics are withheld under the specified minimum-four-cohort support rule. This means insufficient support under that rule, not algebraic nonestimability, an observed pretrend failure or evidence that trends are parallel. Earlier observations or a separately validated age-comparison design could add information.

A survey-window sensitivity ends observations in 2019 while retaining all three weighting rules and proxy-cohort shifts. Its 36 fits are correlated restrictions of the same data, not new studies. At zero shift with equal-cell weighting, youth-obesity ratios are 0.977 (0.960–0.994) for PCV7 and 1.013 (0.977–1.050) for RotaShield; WIC's ratio is 1.014 (0.997–1.031). The PCV7 timing-diagnostic failures remain and WIC's preperiod support remains insufficient. The unshifted asthma sample is unchanged by the cutoff, so its identical result supplies no additional robustness evidence. Excluding later waves does not isolate a pandemic effect or correct other sampling changes. The two rate-analysis scripts accept `--survey-end 2019` and preserve these outputs under separate `_through_2019.json` filenames.

Flagship-index obesity check. The autism flagship's PCV7 index (mean NIS survey-year 2001–2003 coverage with at least one PCV dose; one standard deviation = 8.91 points) correlates 0.805 across 51 states with the chronic-model index above (survey years 2003–2006; one standard deviation = 4.54 points). Under the flagship index, all nine obesity specifications fail the pretrend test (slopes -0.00721 to -0.00498 log points per standard deviation per cohort-year; maximum p = 0.000335) and the 1993 placebo (maximum p = 0.000378). Net of a linear intensity-by-cohort trend, the placebo no longer rejects (p = 0.179–0.897), so the failure is mainly linear drift. The trajectory is not exactly linear, however: net of a whole-window linear trend over cohorts 1985–2008, the post-2000 step is 0.998–1.035, nominally above one in 5 of 9 specifications (for example 1.033, 1.011–1.055), against 0 of 9 under the chronic index (0.982–1.015). Applied to obesity, the flagship design (cohorts 1997–2003, reference 1999, zero shift) gives steps of 1.005–1.007 with fixed effects, 1.016–1.037 with a linear intensity trend and 1.024–1.048 with the spline (smallest p = 0.088), against autism steps of 1.017, 1.016 and 1.013. Its two-lead test rejects on obesity (p = 0.003–0.027) but not on autism (p = 0.636). The autism intensity-by-cohort slope is +0.00029 (95% CI -0.00985 to +0.01043), and the two-lead test has 8.5–12.7% power against drifts the size of the flagship-index obesity pretrend. Injected into the autism fitted means with the fitted step removed, such a linear drift produces a fixed-effects step of 0.975–0.983 instead of 1, while

the linear-trend and spline steps stay at exactly 1.000. In never-eligible cohorts born by 1996, the flagship index does not predict autism at ages 9–17 in IDEA report years 2005–2011 (calendar slope p = 0.831). Treating obesity as a falsification outcome assumes that PCV7 does not affect obesity; this assumption is not tested here. A linear drift is therefore absorbed, but non-linear state-cohort drift producing steps of up to about 1.05 (the flagship-design linear-trend and spline obesity steps are 1.016–1.048) cannot be excluded for the autism flagship (main text Section 4.1). Source: `07_analysis_code/3210_pcv7_intensity_obesity_trajectory.py` and its `.md` and `.csv` outputs.

These exploratory results receive no causal-synthesis weight. Actual vaccine eligibility and uptake, survey design, repeated sampling, proxy-cohort error, programme participation, ascertainment and diagnostic power remain unresolved. Multiple-testing adjustments alone cannot repair these limitations. Reproduction uses `processed/youth_obesity_rate_sensitivity.py` and `processed/asthma_wic_rate_audit.py`; all fitted sensitivities, diagnostics and source hashes are retained in `qa/claude_corrections_20260919/youth_obesity_rate_diagnostics.json` and `asthma_wic_rate_audit.json`. The chronic correction pipeline reproduces these files together with the EURODIAB outputs; this is computational reproduction, not independent causal validation.

S48.1 Non-autism disability and psychiatric panels

These panels relate area-level vaccine coverage or rollout intensity to non-autism disability, psychiatric and physical outcomes. Exposures, outcomes and scales differ by panel. None is an individual vaccinated-versus-unvaccinated risk ratio, and no panel has a policy event with a validated first stage. Rows within a panel share places and cohorts and are correlated. The Norwegian design audit classifies its panel as "exploratory only": municipality and year fixed effects remove stable place differences and common shocks, but exposure is continuous uptake variation without a policy event or credible instrument. The Hong Kong audit records a failed causal gate: national exposure is a function of cohort, and no untreated geography or valid event-study leads exist. All rows receive zero causal-synthesis weight.

Table S48-A. Family-level summaries, counted from every row of each named file.

Panel (file)	Tests	Nominal p < .05	BH q < .05	Placebo and falsification checks
Norway primary-care claims (804): 22 condition-design pairs × 4 timings	88	8 (6 RR > 1, 2 RR < 1)	0 (smallest q 0.394)	0 of 22 future t+1 placebos nominal (smallest p 0.141)
Brazil SIH hospitalizations (888): 11 endpoints × 4 timings × 4 specifications	176	6 (1 RR > 1, 5 RR < 1); all 6 in the unweighted specification, 0 in the primary population-weighted specification	0 (smallest q 0.188; primary lag-0 family 0.555)	Both positive-control families fail to show protection (Table S48-B)
France CNAM descriptive fixed effects (1469)	169 non-placebo, 42 placebo	37 of 169 (8 RR > 1, all intellectual disability at ages 5–9; 29 RR < 1)	2 of 169 non-placebo, both inverse DTP3 (the same 2 under BH across all 211 models); 0 of 42 placebo rows (smallest all-211 q 0.051)	8 of 42 future-younger-cohort placebos nominal. Each of the 8 positive rows passes its matched placebo. At ages 10–14, 5 of 6 intellectual-disability placebos are nominally positive
France CNAM childhood psychiatric, pooled ages 5–9 cohorts (225)	45 non-placebo, 15 placebo	3 of 45 (all RR > 1, all intellectual disability with MMR)	no adjusted column in file	4 of 15 future-cohort placebos nominal: DTP3 0.827, PCV3 0.879 and HEPB3 0.899 for childhood-onset psychiatric conditions, and MMR1 1.492 for other psychiatric conditions; the intellectual-disability MMR2 placebo, 1.100, equals the first-difference estimate

Panel (file)	Tests	Nominal p < .05	BH q < .05	Placebo and falsification checks
Hong Kong special education (910; 909 and 911 for the autism screen)	9 categories	3	3 (two positive, one inverse)	Placebo and false-cutoff checks exist only for autism: future-exposure placebo p = .043; 3 of 10 false cutoffs nominal (2005, 2007, 2013); actual 2009 cutoff p = .807
ACS cognitive difficulty (254–256)	3 PCV7 models (254 rows 0–2)	0 of 3	no adjusted column	Fake first cohorts 1994–2012: 0 of 18 reject with state-year effects (smallest p .362), 5 of 18 with an intensity trend (2001–2005, all RR > 1); 2006 event-study joint pretrend p = .192
Peru CONADIS registrations (896)	80	47 (all RR > 1)	40 of 80	Future-exposure (lag -1) hearing and vision rows are also positive (Table S48-B)
New Jersey IDEA categories (596): 12 categories × 2 scales	24	5	1 (deaf-blindness, rate scale)	—
Taiwan county panel (921)	288	20 (11 RR > 1, 9 RR < 1)	0 (smallest q 0.526)	—
IDEA multivariate, autism + 7 categories (2051)	joint	—	wild-bootstrap familywise inference	Table S48-B

Table S48-B. Selected estimates. Norwegian rows give each nominally significant estimate beside its matched future placebo. Row numbers are 0-based data-row indices in the named file.

Norway: RR per 10 percentage points of municipality coverage. BH q is computed across all 88 tests and equals 0.394 for every nominal row.

Outcome (code group in file)	Coverage; design	Timing	RR (95% CI); p	Matched future t+1 placebo (95% CI); p	Rows (file 804)
Anxiety and depression symptoms (P01_P04ogP25ogP28_P29)	Measles-containing (MESLINGER); child schedule	Contemporaneous	1.131 (1.018–1.255); .021	1.010 (0.865–1.181); .896	17 / 16
Psychological symptoms (P01_P29)	MESLINGER; child schedule	Lag 1	1.093 (1.013–1.179); .022	1.001 (0.934–1.073); .977	22 / 20
Psychological symptoms and disorders (P01_P29ogP70_P99)	MESLINGER; child schedule	Lag 1	1.082 (1.006–1.163); .035	1.015 (0.956–1.077); .633	26 / 24
Musculoskeletal diagnoses, excluding fractures and injuries (L70_L71ogL82_L99)	MESLINGER; child schedule	Lag 2	1.093 (1.006–1.188); .036	1.024 (0.926–1.133); .642	15 / 12
Musculoskeletal complaints (L01_L29)	HPV (HPV_K); adolescent girls	Lag 1	1.007 (1.001–1.014); .034	0.994 (0.981–1.008); .432	50 / 48
Musculoskeletal, excluding fractures and injuries (L01_L29ogL70_L71ogL82_L99)	HPV_K; adolescent girls	Lag 1	1.007 (1.002–1.013); .011	0.999 (0.986–1.012); .865	54 / 52
Affective and depressive disorders (P73ogP76)	HPV_K; adolescent girls	Contemporaneous	0.902 (0.821–0.991); .032	0.906 (0.794–1.033); .141	77 / 76
Anxiety disorders (P74ogP79ogP82)	HPV_K; adolescent girls	Contemporaneous	0.939 (0.887–0.994); .029	1.026 (0.977–1.078); .307	81 / 80

Other panels:

Panel; scale	Outcome; exposure; specification	Estimate (95% CI)	p	q (family)	Role	Row
Brazil SIH; RR per 10 pp coverage, primary raw coverage, population-weighted	Meningococcal infection, infants; meningococcal C; lag 0	1.021 (0.993–1.051)	.147	.975 (176)	Positive control	888: 76
Brazil SIH	Bacterial meningitis, infants; pneumococcal; lag 0	0.961 (0.891–1.035)	.292	.975 (176)	Positive control	888: 68
Brazil SIH	Cerebral palsy, ages 1–4; MMR dose 1; lag 2	0.985 (0.964–1.006)	.150	.975 (176)	Claim-relevant; primary specification	888: 168
Brazil SIH; RR per 10 pp coverage, primary raw coverage, unweighted	Ileus or obstruction, ages 1–4; rotavirus; lag 0	1.0009 (1.0004–1.0015)	.001	.188 (176)	Only nominal row above one; smallest q	888: 67
France (225); RR per 10 pp mean department coverage, pooled ages 5–9 cohorts	Childhood-onset psychiatric; DTP3; department and year FE	0.948 (0.864–1.040)	.260	none in file	Actual exposure	225: 0
France (225)	Same outcome; DTP3; future-cohort coverage placebo	0.827 (0.726–0.941)	.004	none in file	Placebo	225: 3
France (225)	Intellectual disability; MMR1; department and year FE	1.205 (1.010–1.438)	.039	none in file	Actual exposure	225: 32
France (225)	Intellectual disability; MMR2; department and year FE	1.159 (1.023–1.313)	.021	none in file	Actual exposure	225: 36
France (225)	Intellectual disability; MMR2; first difference plus year FE	1.100 (1.006–1.202)	.036	none in file	Actual exposure	225: 38
France (225)	Intellectual disability; MMR2; future-cohort coverage placebo	1.100 (0.964–1.256)	.158	none in file	Placebo	225: 39
France (225)	Other psychiatric conditions; MMR1; future-cohort coverage placebo	1.492 (1.128–1.974)	.005	none in file	Placebo	225: 55
France (1469); RR per 10 pp department coverage	Childhood-onset psychiatric, ages 0–4, girls; DTP3; department and year FE	0.676 (0.589–0.776)	< .001	< .001 (169 non-placebo)	BH survivor	1469: 15
France (1469)	Same; plus department linear trends	0.932 (0.759–1.144)	.499	.759 (169)	Trend version	1469: 16
France (1469)	Childhood-onset psychiatric, ages 0–4, both sexes; DTP3; department and year FE	0.760 (0.666–0.866)	< .001	.003 (169)	BH survivor	1469: 25
France (1469)	Same; plus department linear trends	0.869 (0.692–1.092)	.229	.565 (169)	Trend version	1469: 26
France (1469)	Intellectual disability, ages 5–9, both sexes; MMR1; department and year FE	1.347 (1.056–1.719)	.016	.132 (169)	Nominal positive	1469: 176

Panel; scale	Outcome; exposure; specification	Estimate (95% CI)	p	q (family)	Role	Row
France (1469)	Same; first difference plus year FE	1.352 (1.072–1.707)	.011	.124 (169)	Nominal positive	1469: 178
France (1469)	Same; future-younger-cohort placebo	0.949 (0.784–1.149)	.594	.759 (all 211)	Matched placebo	1469: 179
France (1469)	Intellectual disability, ages 5–9, boys; MMR1; department FE plus linear trends	1.645 (1.031–2.625)	.037	.182 (169)	Nominal positive	1469: 141
France (1469)	Intellectual disability, ages 5–9, boys; MMR1; first difference plus year FE	1.443 (1.121–1.857)	.004	.085 (169)	Smallest q among positive rows	1469: 142
Hong Kong; RR per 10 pp PCV coverage, cohort-aligned, quadratic cohort curve	Intellectual disability	1.033 (1.009–1.057)	.006	.019 (9)	BH survivor	910: 3
Hong Kong	Physical disability	1.049 (1.016–1.084)	.004	.016 (9)	BH survivor	910: 5
Hong Kong	Mental illness	0.937 (0.901–0.974)	.0009	.008 (9)	BH survivor	910: 4
Hong Kong	Autism screen; future exposure (+2 cohort years)	0.990 (0.981–1.000)	.043	none in file	Autism-only temporal placebo	909: 1
ACS; RR per 1 SD rollout intensity at full ages 5–17 eligibility	Cognitive difficulty; state-year FE	0.999 (0.942–1.060)	.975	none in file	Actual exposure	254: 0
ACS	Cognitive difficulty; intensity linear trend	1.082 (0.994–1.177)	.069	none in file	Actual exposure	254: 1
ACS	Cognitive difficulty; future-share placebo	1.000 (0.877–1.141)	.999	none in file	Placebo	254: 2
ACS; RR per 1 SD early PCV uptake	Fake first cohorts 2001, 2002, 2003, 2004, 2005; intensity linear trend	1.097 (1.004–1.198); 1.121 (1.015–1.238); 1.156 (1.016–1.316); 1.188 (1.007–1.402); 1.250 (1.027–1.520)	.041; .025; .029; .042; .027	none in file	Fake-cohort placebos	255: 15, 17, 19, 21, 23
Peru; RR per 10 pp MMR share of EsSalud immunization services, ages 3–5, immunization-volume WLS	Hearing diagnosis; lag 4	1.769 (1.259–2.486)	.001	.005 (80)	Primary-family row	896: 32
Peru	Vision diagnosis; lag 4	1.560 (1.101–2.209)	.012	.031 (80)	Primary-family row	896: 12
Peru	Hearing diagnosis; future exposure (lag -1)	1.416 (1.160–1.729)	< .001	.003 (80)	Future-exposure placebo	896: 20
Peru	Vision diagnosis; future exposure (lag -1)	1.420 (1.163–1.734)	< .001	.003 (80)	Future-exposure placebo	896: 0
Brazil maternal Tdap; RR per 10 pp coverage, any disability, cohort-aligned	Coverage capped at 100, unweighted	1.019 (1.002–1.035)	.024	none in file	Actual exposure	2097: 19
Brazil maternal Tdap	Four weighted coverage variants	1.006–1.013 (all intervals include one)	.089–.247	none in file	Actual exposure	2097: 15–18
Brazil maternal Tdap	Post-birth future exposure (5 rows)	1.000–1.002	≥ .657	none in file	Future-exposure placebo	2097

Panel; scale	Outcome; exposure; specification	Estimate (95% CI)	p	q (family)	Role	Row
IDEA multivariate; 1,999 wild draws, 51 state clusters	Joint null, all eight outcomes' coefficients zero	Wald Q = 19.445	.002	familywise	Omnibus	2051: 12
IDEA multivariate	Largest positive outcome coefficient	max t = 1.061	.606	familywise	Omnibus	2051: 13
SchoolVaxView; RR per 10 points MMR, category share of IDEA disability, ages 6–11	Other health impairment	1.168 (1.013–1.347)	.032	none in file	Claim-relevant category	385: 7
New Jersey; change in rate per 10,000 per 10 pp schedule compliance	Deaf-blindness	-0.729 (-1.139 to -0.319)	< .001	.006 (within family and scale)	Only BH survivor	596: 5
New Jersey; rate ratio per 10 pp schedule compliance	Deaf-blindness	0.287 (0.119–0.690)	.005	.059 (within family and scale)	Same outcome, log scale	596: 4

Notes. The Norwegian condition families overlap. Anxiety and depression symptoms (P01_P04ogP25ogP28_P29) lie within psychological symptoms (P01_P29), which lie within psychological symptoms and disorders (P01_P29ogP70_P99). The two HPV musculoskeletal groups also overlap. The 88 tests are therefore correlated. Six nominal Norwegian rows are in the harmful direction: the three measles-psychiatric-symptom rows and three musculoskeletal rows (measles, lag 2; two HPV rows, lag 1). None of their matched future placebos rejects.

File 888 was regenerated on 24 September 2026 after a source repair; the rows above are from the current file. No primary-specification (population-weighted, raw-coverage) row has $p < .05$. The six nominal rows all come from the unweighted specification: epilepsy at ages 1–4 (lag -1 and lag 0), hemorrhagic disorders at ages 1–4 (lag 0), and ileus or obstruction at ages 1–4 (lags 0, 1 and 2). Only the lag-0 ileus row is above one.

In French file 225, the intellectual-disability MMR2 future placebo, 1.100, equals the first-difference actual-exposure estimate, 1.100; the fixed-effects estimate is 1.159.

The Hong Kong future-exposure placebo and false-cutoff scan were fitted to autism only. Files 909 and 911 contain only autism rows, and file 910 contains only cohort-aligned rows, so no category-specific placebo exists for intellectual disability, physical disability or mental illness. Hong Kong PCV exposure uses official survey coverage of 94.8, 96.2, 96.1 and 96.4 percent for birth cohorts 2009–2012, with earlier cohorts coded zero, so exposure is a single national cohort step. The per-ten-point scale therefore understates the full contrast. Applying the file-910 log coefficients over 9.48–9.64 ten-point units gives implied full-step ratios of 1.355–1.362 for intellectual disability, 1.578–1.590 for physical disability and 0.533–0.539 for mental illness. These contrasts exist only relative to an imposed cohort-curve restriction (quadratic in these file-910 fits; grade and year effects absorb any linear cohort trend; file 913: causal gate failed).

In New Jersey, only the rate-per-10,000 deaf-blindness row survives BH (q .006); the log-rate version has q .059.

The IDEA omnibus rows report the test statistic, not an RR (see also S24). Files 225, 254/255, 385 and 2097 carry no multiplicity adjustment.

Brazilian age-comparison results for the same outcome families are reported in S45.3–S45.4 and are not repeated here. Infant-versus-adult gastrointestinal (A00–A09) admissions, 2007 versus 2005, give a ratio of rate ratios of 0.701 (state-clustered 0.637–0.771; five-region-clustered 0.616–0.798; regional sign-score $p = 0.0625$) (S45.3). In the restricted-support age-period models of S45.4 (the 1996–2017 matched-cell table and the 2000–2017 coverage-adjusted tables), cerebral-palsy (G80) ratios range from 0.768 to 0.939 across samples and denominators, or 0.774–0.886 with the current coverage-adjusted birth denominator (the values Section 4.4 reports). Epilepsy and status epilepticus (G40/G41) range from 0.946 to 1.171, or 0.952–1.079. These are specification ranges, not confidence intervals.

Reproduction: `07_analysis_code/3212_chronic_supplement_tables.py` reads each row from the file and row key shown and writes `3212_chronic_supplement_tables.md`, `3212_nonautism_panel_results.csv` (SHA-256 `3b0ebe1a9cd396134d2d2ea3aff4b4496c59aab5ec85acf7913cf5effd10b036`) and `3212_eurodiab_introduction_family.csv` (SHA-256 `7f8ab4f78ecc09773ad50420949daab98ffb92183dd22bcce8d533c863aae6fc`). Input-file hashes are listed in the

Markdown output. The Hong Kong implied full-step ratios, the Brazil unweighted-row listing and the file-225 and Peru nominal counts are computed directly from files 910, 888, 225 and 896.

S49. External diabetes evidence and population dependence

Published rotavirus/type 1 diabetes associations are not uniformly null. Full-series versus unvaccinated adjusted HRs include 1.03 (95% CI 0.62–1.72) in Glanz et al., with a partial-series HR of 1.50 (0.81–2.77); 1.09 (0.87–1.36) in Burke et al.; and 0.67 (0.54–0.83) in Rogers, Basu, and Kim, or 0.70 (0.56–0.88) with birth year.

These selected external studies are context, not an exhaustive review or constituents of our causal synthesis. VSD, MarketScan and Optum are distinct data systems, but distinct systems do not establish participant independence. Shared controls, alternative specifications and age/product subsets also cannot be counted as independent replications. Enrollment requirements, follow-up and confounding control differ. Neither the inverse association establishes prevention nor the near-null estimates individually exclude small or moderate effects. Outcome-specific appraisal is necessary; autism findings cannot establish a universal chronic-disease null.

S49.1 Randomized, birth-date and published person-level chronic-outcome evidence

Type 1 diabetes. Three comparisons allocated Hib or rotavirus vaccination by randomization or birth date; none supports a precision or exclusion claim. The Finnish Hib conjugate trial assigned vaccination at 3, 4, 6 and 14-18 months or at 24 months by odd or even birth date (Eskola et al. 1990). Over ten years diabetes risk did not differ significantly between arms, RR 1.06, a contrast of four early doses with one dose at 24 months rather than vaccination versus none; the delayed arm's non-randomized comparison with an older pre-trial cohort gave RR 1.01 (Karvonen, Cepaitis, and Tuomilehto 1999; no intervals extracted). Reanalyzing the same trial, Classen and Classen (2002) compared the four-dose arm with that pre-trial cohort and reported a cumulative-incidence difference of 54 cases per 100,000 at seven years (relative risk 1.26, p = 0.026), mostly in clusters beginning about 38 months after immunization that they located by searching the follow-up data. That harm claim rests on a non-randomized comparison across birth years and on data-derived windows, so, like Karvonen's historical contrast, it receives zero weight. A follow-up of the Northern California HbOC trial cohort to ages 10-12, comparing children offered the vaccine with children excluded by birth date, reported no evidence of association, with no ratio in its abstract (Black et al. 2002).

In the randomized Finnish Rotavirus Efficacy and Safety Trial extension, parents of 5,764 of 19,133 participants (30 percent) returned a 2015 questionnaire; type 1 diabetes was reported for 25 of 2,580 placebo and 33 of 3,184 RotaTeq recipients (p = 0.810), and celiac disease for 1.11 versus 0.60 percent (p = 0.027) (Hemming-Harlo et al. 2019). Unequal respondent arms under 1:1 allocation suggest differential response, so neither comparison inherits randomization, bounds harm or shows protection.

In a Norwegian natural experiment covering 740,744 children born 2007-2019, with 846 type 1 diabetes cases between six months and five years, 92.8 percent of children born on or after the 1 September 2014 rotavirus cutoff were fully vaccinated, against 0.3 percent of those born before it (Östman et al. 2026b). An interrupted-time-series Cox model estimated a harmful-direction change in the log-linear birth-cohort trend at the cutoff, relative HR 1.11 (1.03-1.20), or 1.05 (0.93-1.18) excluding the 2013-2015 cohorts. This is a slope change, whose time unit we did not extract, not a level risk ratio. Rotavirus specificity is not demonstrated: post hoc breaks at the hepatitis B introduction and the PCV7-to-PCV13 switch have nearly the same point estimates, 1.10 (0.96-1.28) and 1.10 (0.97-1.24), though their intervals include one, and because those cutoffs are also vaccine changes they cannot distinguish secular trend from a schedule-wide effect. No pretrend test was reported. The adjusted HR for children fully vaccinated against rotavirus versus those not vaccinated against it, 1.17 (0.78-1.77), is observational. In the same register cohort, celiac disease showed no significant trend break at the cutoff (p = 0.46); the born-after-versus-before HR was 0.96 (0.89-1.04), and the unadjusted observational HR for fully vaccinated versus unvaccinated children was 0.99 (0.92-1.07); the adjusted Table II estimate is 1.02 (0.82-1.28) (Östman et al. 2026a). That study shares its cohort with the diabetes study and is not independent of it. Both designs receive zero causal weight.

Asthma. A Swedish register follow-up compared almost 80,000 children vaccinated against pertussis in a 1993-1994 trial with about 100,000 children born in the months just before or after it, who received no pertussis vaccine; after 14 years without general pertussis vaccination, the routine schedule gave diphtheria and tetanus toxoids without pertussis. Dispensed asthma medication at age 15 had adjusted OR 0.99 (0.95-1.03) both by birth period (intention to treat) and for trial participants (per protocol) (Vogt et al. 2014). This is a pertussis-component contrast, not vaccination versus none,

and without timing diagnostics it is triangulation with zero weight. A Vaccine Safety Datalink cohort of 326,991 children found persistent asthma at 24-59 months positively associated with aluminum before age two, adjusted HR per milligram 1.19 (1.14-1.25) without eczema and 1.26 (1.07-1.49) with it (Daley et al. 2023), against a Danish HR of 0.96 (0.94-0.98). That observational association, which its authors note may reflect residual confounding, receives zero weight but is not dismissed because the Danish estimate is inverse.

Aluminum and fifty chronic outcomes. Andersson et al. (2025) related cumulative vaccine aluminum by age two to 50 chronic outcomes at ages two to five among 1,224,176 Danish children. Per milligram, HRs were 0.93 (0.90-0.97) for any neurodevelopmental disorder, 0.93 (0.89-0.97) for the autism spectrum disorder composite, 0.90 (0.84-0.96) for ADHD, 0.98 (0.94-1.02) for autoimmune disorders, 0.99 (0.98-1.01) for atopic or allergic disorders and 1.03 (0.92-1.15) for type 1 diabetes. Among separately analyzed outcomes, upper limits exceed 1.30 for rheumatoid arthritis 1.10 (0.72-1.68), Hashimoto thyroiditis 0.92 (0.54-1.56), idiopathic thrombocytopenic purpura 1.13 (0.97-1.31), insect-sting allergy 1.09 (0.66-1.79) and Asperger syndrome 1.13 (0.89-1.44); atopic dermatitis, 1.02 (1.00-1.04) on 22,978 cases, is the only individual main-text outcome whose printed lower limit is 1.00 rather than below one. The study is observational and contrasts dose, not vaccination versus none; its inverse HRs are not evidence of prevention. A correction two days after publication replaced "an incorrect version of the Supplementary Material" (Annals of Internal Medicine 2025). Main-text estimates did not change, but supplementary neurodevelopmental analyses did: the sex-stratified case count rose from 2,239 to 5,200, and the corrected supplement names the Psychiatric Central Register as the outcome source, which the Methods do not (they name the National Patient Registry and, for ADHD, prescription fills). The corrected Asperger estimate among 2007-2018 births is 1.67 (1.01-2.77) on 51 events, without multiplicity correction, against 0.75 (0.41-1.40) on 17 events originally; at age eight, Asperger syndrome is 1.02 (0.93-1.12). For chronic outcomes other than autism, the observational aluminum evidence is mixed.

Epilepsy. All person-level evidence located is observational. The one study of both vaccination status and vaccine aluminum, a Vaccine Safety Datalink case-control study of incident epilepsy before age four (2,089 cases, 20,139 matched controls), reports that no adjusted odds ratio in the primary analysis exceeded one for up-to-date status or for cumulative aluminum per milligram per adjuvant formulation, and that analyses by age group and restricted to epilepsy of unknown etiology showed no statistically significant relationship (McClure et al. 2026). Its odds ratios were not accessed, and point estimates without intervals supply no upper bound on harm. The authors' preliminary 2024 poster estimates (19,770 children) are not the final article's results and are neither quoted nor plotted. Cohort HRs are 1.01 (0.66-1.56) for epilepsy later in childhood after Danish DTaP-IPV-Hib vaccination (Sun et al. 2012) and 1.07 (0.94-1.23) after Norwegian pandemic influenza vaccination (Håberg et al. 2018).

Unadjusted chronic-harm contrasts. In Florida Medicaid claims for 47,155 nine-year-olds, children with billed vaccination visits had any neurodevelopmental disorder at OR 3.12 (2.85-3.41) relative to children without one (Mawson and Jacob 2025). An unpublished Henry Ford Health draft entered into a 9 September 2025 Senate hearing record reports an adjusted hazard ratio of 2.54 (2.16-2.97) (its abstract gives 2.53, 2.16-2.96) for chronic conditions among 16,511 vaccinated versus 1,957 unvaccinated children, noting that unvaccinated children use less healthcare (Lamerato et al. 2025). Such contrasts are open to healthcare-utilization and ascertainment bias and receive zero weight. These external studies are context for the chronic register (Figures S69-S71), not constituents of the causal synthesis.

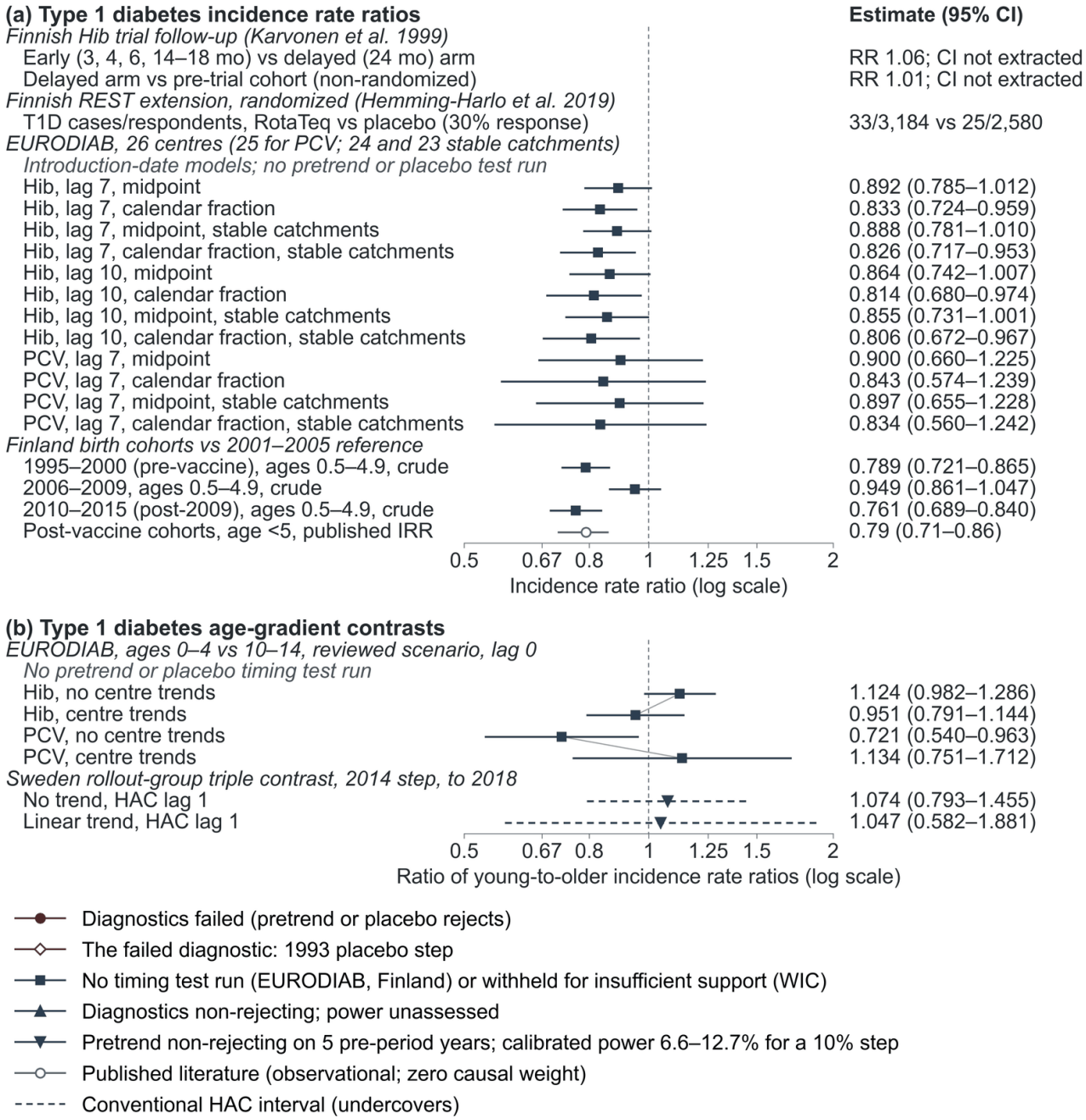


Figure S69. Chronic-outcome register, panels a-b. Native diabetes incidence ratios, introduction-date and age-gradient contrasts are shown separately. Finnish trial rows without extracted intervals are text-only. Stable catchments excludes catchments with recorded boundary changes. EURODIAB specifications reuse populations; connectors link trend specifications, not independent studies. Swedish conventional HAC intervals can undercover in this short series. No pooled effect is estimated. See S47-S50 for timing, denominator and power limitations.

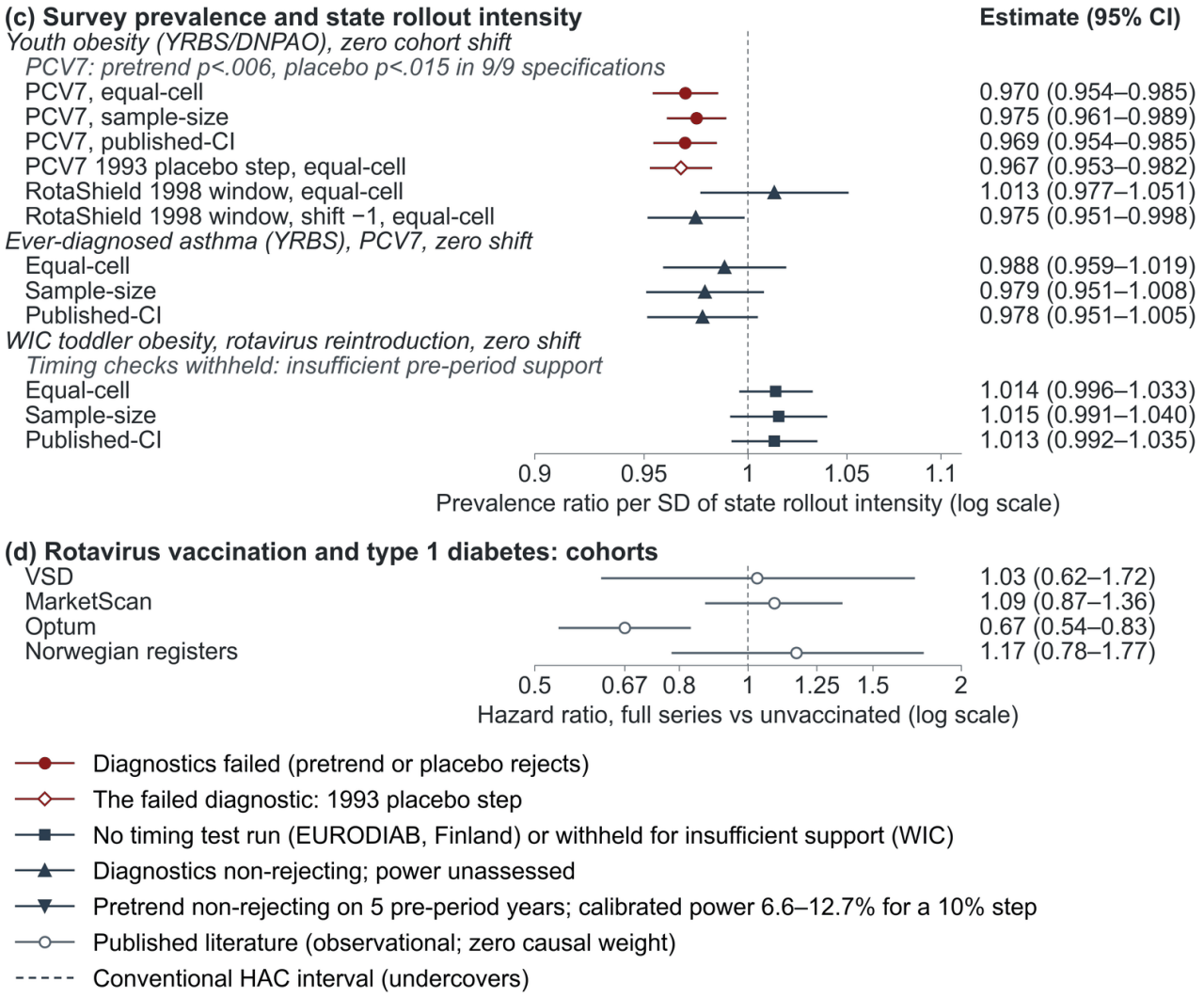


Figure S70. Chronic-outcome register, panels c-d. Survey prevalence ratios are per SD of the stated state rollout index; the PCV index differs from the autism flagship. Published rotavirus cohorts report hazard ratios for full receipt versus no receipt. These scales and source populations are not pooled. Observational estimates have zero causal weight in the autism synthesis. Diagnostics and exposure units remain source-specific.

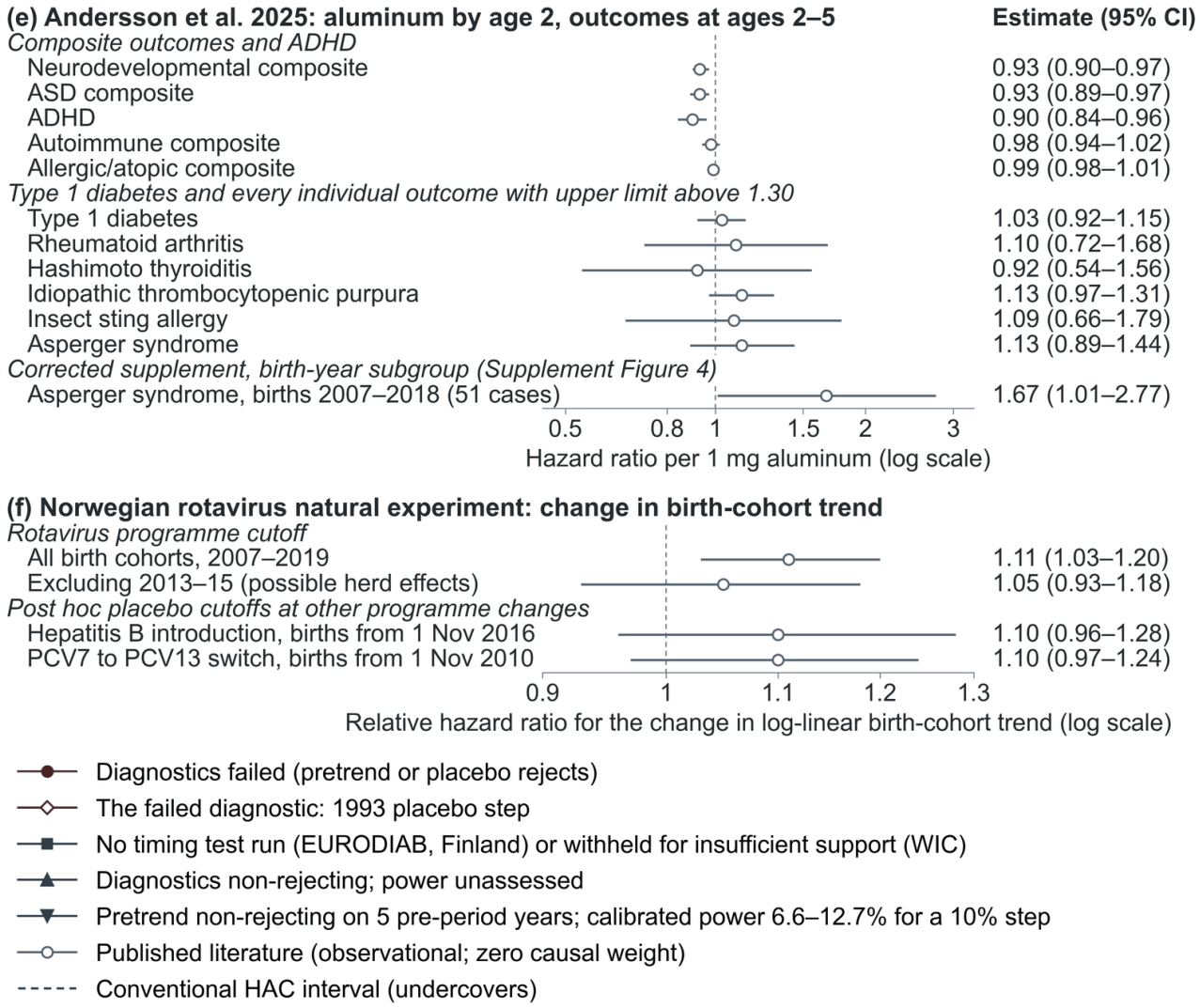


Figure S71. Chronic-outcome register, panels e-f. Danish hazard ratios are per milligram of aluminum by age two; the corrected Asperger subgroup contains 51 cases and has no multiplicity correction. Norwegian estimates compare birth-cohort log-trends at program and placebo cutoffs, not individual receipt risks; the slope time unit was not extracted. The rotavirus cutoff ratio 1.11 is accompanied by placebo-cutoff breaks of nearly the same size (1.10 at both the hepatitis B introduction and the PCV7-to-PCV13 switch; intervals include 1), so the break is not rotavirus-specific; no pretrend test was reported. Positive as well as inverse estimates are retained; none enters a common disability effect.

S50. Swedish annual diabetes contrasts and small-sample power

The appendix to Rangert et al. (2023) supplies annual cases and population denominators by rollout group and age, 2009–2019. These are incident diabetes counts, distinct from annual specialist-care patient counts. Source reconciliation identifies inconsistent middle-age population denominators in the late-adopting group; those cells are excluded without altering the published source. National and all-age totals are overlapping summaries, not additional observations.

We construct annual log incidence ratios between early and late rollout groups for ages 0–4 and 10–14, then difference these age contrasts. OLS compares post-2014 step and five-year ramp specifications, with and without linear trend adjustment, endpoints of 2018 or 2019, and Bartlett HAC lags 0–2. The 2014 transition year is omitted. The ramp is a cohort-replacement proxy, not measured uptake. All 72 exploratory specifications are retained. Collapsed rollout groups do not permit region-clustered inference.

For the step specification ending in 2018 with HAC lag 1, the age-by-region contrast is 1.074 without trend adjustment and 1.047 with it. Conventional t intervals are 0.793–1.455 and 0.582–1.881, respectively, but the simulations below show that nominal coverage is unreliable. Five-year preperiod slope tests have limited power; the

older-age regional contrast has an unadjusted lag-1 p value of .045. Neither a nonsignificant coefficient nor a nonsignificant pretrend provides a precise exclusion bound.

The component series in the same specification (step, end 2018, HAC lag 1) show where the variation lies. The ages 0–4 regional contrast is 0.929 (0.694–1.244) without trend adjustment and 0.704 (0.445–1.114) with it. The ages 10–14 regional contrast is 0.865 (0.725–1.033) without trend adjustment and 0.672 (0.481–0.939; p = .027) with it. Children aged 10–14 after 2014 were born before rotavirus vaccination, so this older-age change cannot reflect their own vaccination. Across all 72 fits, eight have p < .05, and all eight are older-age regional contrasts. Two end in 2018 (step with trend, HAC lags 1 and 2: 0.672). Six end in 2019: three steps (0.731–0.849) and the unadjusted ramp at all three HAC lags (0.805). No young-age or age-differenced fit is nominally significant. Pre-2014 slope tests use five years and three residual degrees of freedom. At HAC lag 1 they give p = .188 for the young contrast, .045 for the older contrast and .941 for the age-differenced contrast; the older-contrast p is .090 at lag 0 and .042 at lag 2.

Monte Carlo calibration evaluates correctly specified aggregate steps under Gaussian AR(1) errors with rho 0, .3, .6 or .9 and two noise scales derived from the five preperiod years. Each covariance/design scenario uses 5,000 independent null-training draws and 5,000 evaluation draws. Nominal 5% HAC tests reject true nulls in 4.8–17.1% of configurations' evaluation samples at rho=0 and 14.8–50.0% at rho=.9. These conditional scenarios demonstrate potential undercoverage, not Sweden's actual false-positive rate.

For the reference configuration used in the main text, the conventional HAC test performs as follows. The configuration is the age-differenced contrast, 2014 step, analysis ending 2018, HAC lag 1, with the noise scale taken from the raw preperiod SD ($\sigma = 0.1625$). Across the eight AR(1)-by-trend scenarios, nominal 5% tests reject a true null in 9.8–44.5% of evaluation samples (0.0976–0.4448): 9.8–44.5% without trend adjustment and 13.1–21.0% with it. Conventional 95% interval coverage is 55.5–90.2%. Critical values calibrated on independent null draws ($|t|$ = 2.87–6.80) restore null rejection to 4.8–5.6%.

AR rho	Linear trend	Conventional null rejection	Calibrated null rejection	Calibrated power: 10% step	25% step	50% step
0	No	9.8%	5.6%	12.7%	40.6%	84.6%
0	Yes	13.1%	5.4%	6.7%	11.3%	25.5%
.3	No	18.0%	5.6%	9.8%	28.0%	67.7%
.3	Yes	17.5%	5.6%	6.8%	10.9%	22.8%
.6	No	28.9%	5.2%	7.9%	21.7%	54.5%
.6	Yes	21.0%	5.0%	6.6%	11.3%	25.3%
.9	No	44.5%	5.6%	11.2%	36.0%	75.9%
.9	Yes	19.0%	4.8%	8.8%	26.7%	62.2%

Each cell uses 5,000 evaluation draws (Monte Carlo SE about 0.3 percentage points at 5%). Power figures are for aggregate step ratios of 1.10, 1.25 and 1.50, not individual vaccine effects.

For the age-differenced contrast ending in 2018, HAC lag 1 and raw preperiod SD, independently calibrated tests have approximately 5% size within each assumed scenario. Across the four rho values and trend choices, power is 6.6–12.7% for a 10% aggregate step, 10.9–40.6% for a 25% step and 22.8–84.6% for a 50% step. These are not individual-vaccine-effect power calculations, and calibration does not identify the true covariance or remove confounding. This module receives no causal-synthesis weight.

Reproduction: `processed/acquire_sweden_t1d_appendix.py`, `audit_sweden_t1d_appendix.py`, `sweden_t1d_annual_contrasts.py`, `test_sweden_t1d_annual_contrasts.py`, and `sweden_t1d_simulation.py`. The observed-rate SVG is produced by `plot_sweden_t1d_annual_contrasts.py`. Complete outputs, source hashes and covariance scenarios are retained in the chronic-outcome audit directory.

S51. Country diabetes sources, timing and data availability

The eight-country compilation (doi:10.1016/j.focus.2025.100409) assembles earlier publications rather than newly recruited independent cohorts. It cannot be pooled alongside its constituent studies as additional participants. Our primary-record inventory covers ten DOI-matched records, including an Australian correction and two Finnish publications; bibliographic coverage is not equivalent to complete annual-data acquisition or risk-of-bias appraisal. The compilation draws on nine earlier publications describing eight countries: Australia, Austria, the Dominican Republic, Finland (two publications), Ireland, Israel, Mexico and Wales. Our ten records are these nine plus the Australian correction. It excluded one Swedish study because that study had no national data for the post-vaccination period (after September 2019). The compilation lists Mexico's introduction as May 2007. Yet its summary of the Mexican under-five series shows incidence rising 47.5% a year (95% CI 11.1–96.0%) in 2001–2005 and falling 20.8% a year (-24.9% to -16.5%) from 2005 to 2018. The decline therefore begins before the listed introduction and cannot be read as a post-introduction response. The Mexican primary study (doi:10.1016/j.diabres.2019.107949) covers children and adolescents insured by the Mexican Institute of Social Security, not all Mexican children. Its abstract also places the turning point before 2007: for everyone under 20, incidence rose from 2000 to 2006 and fell from 2006 to 2018 (annual percentage changes +16.1 and -8.7). The Australian constituent study (doi:10.1001/jamapediatrics.2018.4578) was corrected after a case-identification coding error (doi:10.1001/jamapediatrics.2019.2523). The age 0–4 rate ratio after introduction changed from 0.86 (0.74–0.99) to 0.85 (0.75–0.97), and the study's overall mean rate from 12.7 to 24.4 per 100,000. The compilation reports the corrected 0.85 (0.75–0.97). Its sole author is also the first author of the Optum cohort study in S49, which the compilation cites.

Austria's supplementary Table S1 (doi:10.1111/pedi.13038) supplies annual age-standardized incidence and standard errors for ages 0–14. The combined, male and female T1D series each cover 1989–2017; T2D covers 1999–2017. The extraction preserves 106 observed rate/SE pairs and ten uncollected pre-1999 T2D slots as missing. An independent parser reproduces all 116 slots. These are overlapping outcome/sex summaries, not 106 independent populations. T2D has large relative standard errors, 33.8–100%; this is descriptive precision, not vaccine-effect power. The table does not supply under-five annual rates or raw cases. Reported data-selected joinpoints are not fixed-date vaccine effects, and sales-based uptake is not individual receipt. Type 2 diabetes is a substantive outcome, not an assumed negative control. Reproduction uses `processed/acquire_austria_diabetes_source.py`, `extract_austria_diabetes_rates.py` and `test_austria_diabetes_extraction.py`.

The Welsh primary study (doi:10.1016/j.diabres.2021.108739), Table 2, places fitted under-five incidence peaks in March 2006 for girls and April 2006 for boys, preceding the 2013 rotavirus programme. They are quadratic-model extrema without reported peak-date intervals, not vaccine-effect estimates. Age-specific trend differences mean older children are not automatically valid controls. Annual rates are plotted, but exact annual cells have not been acquired for our reanalysis. Ascertainment corrections, counterfactual curvature, cohort replacement and sampling uncertainty must be addressed before attributing any post-introduction deviation to vaccination.

These sources add information on outcomes and design feasibility, not a new pooled causal estimate. Shared national registers, overlapping observation periods and EURODIAB participation require reconciliation before independent synthesis weights can be assigned. Public availability of an article does not establish permission to redistribute its files; retrieval links and provenance are retained separately from release permissions.

Finland's publisher-deposited supplement, Table S1, supplies observed cases and person-years for four birth-cohort groups through 2016. Eleven nonoverlapping combined-sex age/cohort cells sum to 8,674 cases. The full 45-cell extraction also retains sex-specific and all-age summaries, which overlap those eleven cells and do not add participants. No oldest-age cell is observed for the latest cohort group. Five printed rates lie just outside simple one-decimal rounding of cases divided by person-years; original values and calculated rates are retained separately, without changing source counts.

For ages 0.5–4.9, crude rate ratios relative to the 2001–2005 birth cohorts are 0.789 (Poisson-Wald 95% interval 0.721–0.865) for 1995–2000, 0.949 (0.861–1.047) for 2006–2009 and 0.761 (0.689–0.840) for 2010–2015. These are descriptive count-based reconstructions, not adjusted vaccine-effect estimates. The lower rate in an earlier prevaccine cohort demonstrates why a later-versus-reference reduction alone does not isolate vaccination. The published adjusted incidence rate ratio for the postvaccine versus the prevaccine 2001–2005 birth cohorts under age five is 0.79 (95% CI 0.71–0.86) in the abstract of Parviainen et al. (2024) and 0.79 (0.71–0.87) in the publisher version's Table 2, whose crude ratio, 0.76 (0.69–0.84), matches the reconstruction above. Shared-reference dependence, changing attained-age

composition and secular influences remain. Published sensitivity analyses use predicted cases to augment incomplete age follow-up; predicted cells are not additional observed outcomes. Independent reproduction of the adjusted model and reconciliation with other Finnish register studies remain necessary. Scripts: `processed/acquire_finland_t1d_supplement.py`, `finland_t1d_grouped_counts.py` and `extract_finland_t1d_table.py`.

Sex-specific exploration retains 24 crude cohort ratios and eight direct male-to-female ratios of cohort rate ratios, with separate Holm adjustments for the two families. Interactions use the difference of log rate ratios and the sum of their four inverse case counts as a Poisson variance approximation. They are descriptive sex contrasts, not vaccine-by-vaccine interactions or adjustment for exact age and calendar period. Among ages 0.5–4.9, the 2010–2015-versus-2001–2005 male/female interaction ratio is 0.912 (95% interval 0.748–1.112). One of eight interactions is nominally significant: ages 10.0–14.9 in the 2006–2009 cohort group, ratio 0.355 (0.169–0.747), Holm-adjusted p approximately .051. That cohort/age cell contains only 11 male and 21 female cases with incomplete age follow-up. It is retained as an imprecise descriptive difference, not a vaccine-specific subgroup finding; failure to cross an adjusted threshold does not establish equivalence. Scripts `processed/finland_t1d_sex_contrasts.py` and `test_finland_t1d_sex_contrasts.py` retain all results and test known-ratio calculation, row-order invariance, exclusion of overlapping all-age summaries and sex-swap symmetry.

S52. PCV7 reference exposure and estimand validation

The rollout-intensity spline compares cohorts 2000-2003 with cohorts 1997-1999 across states ranked by early uptake. The ranking averages NIS survey years 2001-2003; one standard deviation is 8.91 percentage points in that index. It is not the residualized cohort exposure difference. Under the survey-year-minus-two proxy, cohort 1999 already has mean recorded pneumococcal receipt of 34.4%; the birth-window remap gives 45.0%. Routine PCV7 vaccination began in 2000, but catch-up eligibility extended beyond infancy. Contemporary ACIP recommendations included high-risk children aged 24-59 months and consideration of other children in that age range (CDC, 2000). Thus 1997 is an assumed low-exposure reference, not an observed unvaccinated cohort.

The NIS field P_NUMPCV records pneumococcal-containing doses, not exclusively marked conjugate doses; DPCV1 records age in days at the first pneumococcal-containing dose. Any-dose and infant-first-dose definitions must be distinguished. The proxy cohort assignment and the remap are alternative approximations, not exact birth-date linkage. The remap assumes uniform interview timing and ages within published AGEGRP bands; the 1998 and 1999 birth windows are incompletely observed. These limitations prevent interpreting the exposure series as verified individual receipt histories for IDEA children.

S52.1 First stages and exposure-aligned injections

With the published spline controls, the first stage regresses recorded coverage on post-2000 times the state intensity ranking. Its any-dose coefficient is negative under both cohort assignments. Reproducing the outcome coefficient does not validate this first stage.

Exposure definition	First stage, percentage points per intensity SD	Returned rollout RR when individual RR 2 is injected
Any pneumococcal dose, proxy cohort	-1.372	0.9625
Any pneumococcal dose, remapped cohort	-6.690	0.9395
First pneumococcal dose before seven months, remapped cohort	6.217	1.0308

The injections use noise-free counts equal to a state-and-cohort fixed-effects Poisson mean multiplied by 1 + (R - 1)D, where D is the stipulated recorded exposure share and R the injected homogeneous individual risk ratio. For R equal to two, three, five and ten, nine of twelve spline results fall below one. This establishes a failure of monotone interpretation under those exposure scenarios; it does not estimate true vaccine harm, sampling power or the probability that the assumed exposure reconstruction is correct. An independently specified matrix regression reproduces all twelve results within 1e-8, conditional on the shared upstream exposure reconstruction.

S52.2 Alternative reference and inference

Referencing the event study to 1997 gives an average 2000-2003 RR of 1.025 (state-cluster CR1 95% interval 0.982-1.070; 50 degrees of freedom), reproduced by a separate matrix fit. It is a contrast per intensity standard deviation. Alternative coverage-rescaled and minimum-distance estimates in the audit retain uncertainties about reference catch-up, exposure alignment and proportionality. They use the same data and are not additional independent evidence. Selection of a reference because its injection results have the expected sign does not establish causal identification.

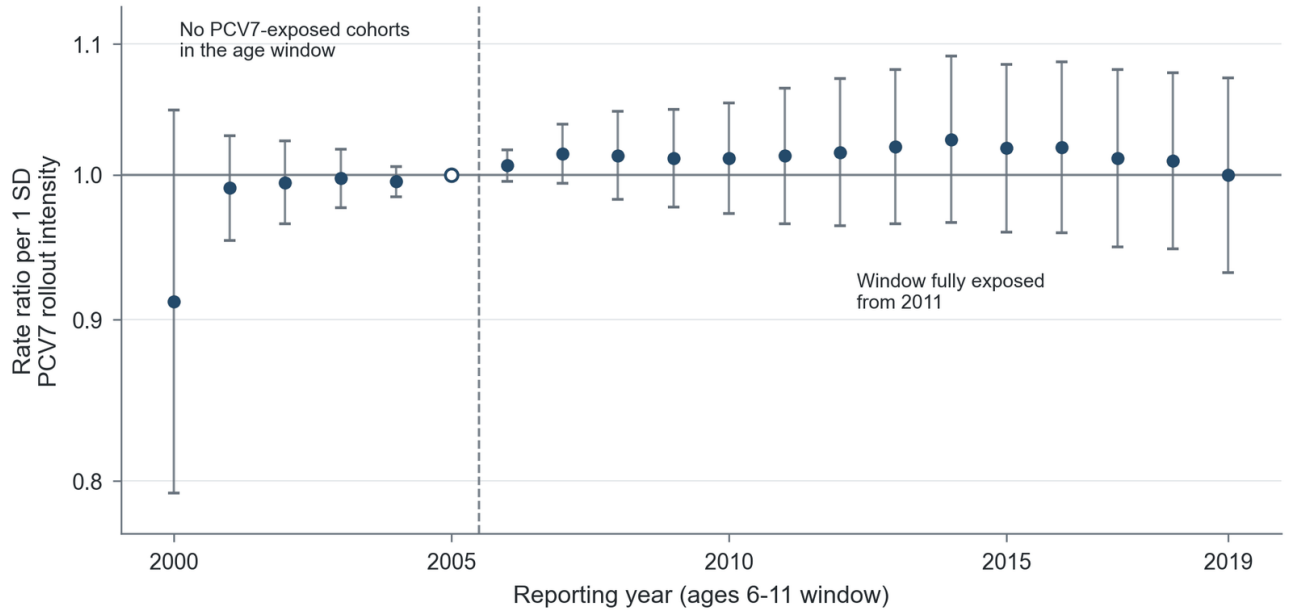


Figure S73. Cohort contrasts by early PCV7 rollout ranking. The five historical pre-exposure coefficients span an outcome-source transition at the reference year and are a sensitivity display, not five clean tests of a single unchanged measurement regime. This outcome-only display does not establish unexposed reference cohorts or validate an individual-risk mapping.

S52.3 Scope of numerical calibration exercises

The PCV7 model-native MDEs, homogeneous-mixture individual-equivalent calculations, numerical posterior integration checks and cross-design sign-synthesis calculations elsewhere in this supplement do not establish individual-risk exclusion. The first-stage audit invalidates treating the spline coefficient as the high-minus-low coverage contrast required by that mixture equation. Numerical agreement of integrators, source-to-table reproduction and sensitivity to prior masses cannot repair this mismatch. Those algorithmic calculations are not used as vaccine-risk posteriors or as a universal null probability. RotaShield and every other family require their own exposure mapping and identification assessment; this audit neither validates nor invalidates them automatically.

S52.4 Pneumococcal product-definition sensitivity

A separate parser reads the original NIS 2001-2006 ZIP members using their public SAS layouts. All dose totals reconcile as marked conjugate plus unknown-type plus marked polysaccharide doses in the 126,856 adequate-provider records. In survey year 2001, national weighted any-dose receipt is 37.05%, compared with 30.47% marked conjugate receipt; 5.62% have an unknown-type dose but no marked conjugate dose, and 0.95% have only marked polysaccharide doses. These are survey-year weighted percentages, not the state-unweighted proxy-cohort means above. Unknown-type receipt is not assumed unvaccinated.

On the same proxy-cohort outcome panel, six executions compare three receipt definitions under the original all-pneumococcal rollout ranking and a ranking recomputed for each definition. The all-pneumococcal baseline repeats across rankings and adds no information.

Receipt definition	Ranking	Spline first stage, pp per SD	State-cluster SE
All pneumococcal	Original or matched	-1.372	1.604
Marked conjugate	Original	-0.338	1.442
Conjugate or unknown type	Original	-1.339	1.524

Receipt definition	Ranking	Spline first stage, pp per SD	State-cluster SE
Marked conjugate	Definition-matched	0.121	1.390
Conjugate or unknown type	Definition-matched	-1.162	1.501

The corresponding 24 noise-free injections at individual RRs of two, three, five and ten all return rollout ratios below one; four duplicate the baseline execution. These are deterministic sensitivity calculations, not independent observations or calibrated sampling-power tests. Excluding marked polysaccharide-only receipt does not resolve the mapping failure in these scenarios. Type recording, unobserved catch-up, within-age-band timing and the assumed zero exposure of earlier proxy cohorts remain limitations. Identified conjugate receipt is not a complete measure of true conjugate receipt when type is unrecorded.

Sources are the original NIS public-use records and layouts; outputs 3200-3204 document the cohort audit. The companion reproduction guide identifies the source-definition parser, independent matrix refits and product-definition sensitivity scripts. Run the held-input checks with `python processed/run_pcv_exposure_review.py`. Each execution creates dated receipts with input and code hashes. Original coefficients and simulated results are retained. Computational agreement is distinct from scientific identification.

S53. Design-matched individual-scale inference

S53.1 Identification and conversion are separate tasks

Ecological measurement is not a categorical bar to causal inference. A valid policy contrast, an aligned receipt first stage and defensible exclusion assumptions can identify a local average effect among people whose receipt changes. For additive outcomes, the ratio of policy outcome and receipt effects targets a complier risk difference under the relevant IV assumptions, not automatically a population RR. Weak first stages require confidence-set inversion with the joint covariance, not division of independently estimated interval endpoints (Angrist, Imbens and Rubin, 1996; von Luxburg and Franz, 2009). Ecological confounding and aggregation requirements remain model-specific (Greenland and Robins, 1994).

An alternative is a direct risk-mixture mean: expected cases equal population times counterfactual baseline risk times [1+p(R-1)]. Here p must describe the exposure of the population producing the outcomes. With heterogeneous baseline risk and effects, the increment is E[D Y(0)(R_i-1)], not generally p times E[Y(0)] times E[R_i-1]. A homogeneous multiplier and appropriate baseline-risk composition are sufficient conditions; a correctly specified heterogeneous model can replace them. A scalar fitted R need not be the person-weighted average effect or bound every subgroup.

The executable conversion contract covers all fourteen integrated aggregate/policy rows, twenty-three individual-study rows, and four additional headline families. It preserves native effect measures, overlap blocks, conversion requirements and uncertainty status. Individual HR, OR and RR entries are not automatically interchangeable. Kinks require an exposure trajectory and the same temporal operator; age-gap and DDD contrasts require exposure effects in both comparison groups; schedule interactions require joint exposure and main effects. These are requirements for constructing additional conversions, not declarations that such conversions are impossible.

S53.2 PCV7 direct mixture and signed rollout-score inversion

The direct model fits log(rate)=Xb+log[1+p(R-1)] by nonlinear least squares, profiling the linear nuisance terms. Four baselines are crossed with three exposure maps and specified catch-up scenarios, giving twenty-eight specifications. State-cluster CR1 score tests with t(50) critical values yield nominal acceptance sets; they condition on fixed reconstructed exposure.

The second approach preserves the original rollout regressor as the test direction. For each candidate R, a Poisson model fits nuisance coefficients with log population plus log[1+p(R-1)] as an offset. The rollout regressor is residualized against nuisance covariates using fitted Poisson weights. Its cluster-summed residual score is studentized using state-cluster CR1 variance and t(50). Inversion does not divide by a first stage or require a positive exposure loading. The rollout regressor is not asserted to be a valid instrument. Full conditional-mean restrictions remain load-bearing.

Baseline, remapped any-dose exposure	Direct fitted R	Direct nominal 95% set	Rollout-score nominal 95% set
State and cohort effects	1.446	0.976-2.088	0.164-2.190
Uptake spline	1.727	1.084-2.642	0.017-1.367
State-specific linear trends	0.976	0.543-1.462	0.308-2.706
State trends plus uptake spline	1.008	0.485-1.583	lower search edge-1.486

These rows assume zero otherwise unmeasured 1997 receipt. The spline score equation has roots at approximately R=0.149 and 0.837, not one uniquely identified point estimate. Allowing half or all of the state's 1999 receipt in 1997 gives sets 0.078-1.642 and 0.258-2.275. Some infant-exposure specifications have disconnected sets or upper components reaching R=10,000. The complete archive retains all detected components and boundary flags. A finite search endpoint is not a finite identification bound; a grid search cannot exclude an arbitrarily narrow undetected component.

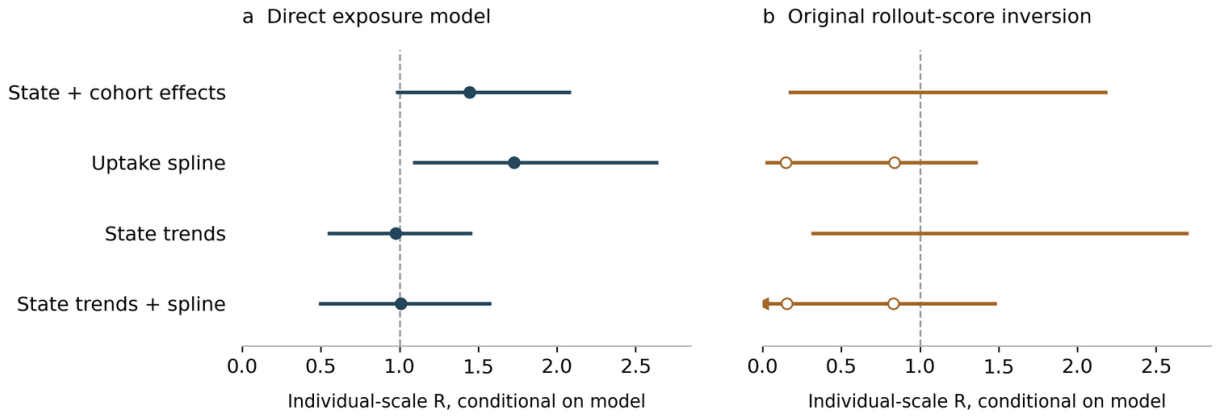


Figure S55. Conditional individual-scale PCV analyses. Same remapped receipt, zero unobserved 1997 catch-up, and the same 353 state-cohort cells. Filled points are direct-model estimates; open points are rollout-score roots, which need not be unique. Lines are nominal state-cluster acceptance sets; the arrow marks the lower search boundary. These are fixed-exposure model comparisons, not independent studies or unconditional causal bounds.

Figure note: both panels use the same 353 state-cohort observations and fixed remapped exposure. Filled points are direct-model fits; open points are roots of the rollout-score equation, not independent or uniquely identified estimates. Lines are nominal acceptance sets. The left-pointing marker denotes a lower search boundary. Different moments and weighting can respond differently when the structural mean is imperfect. Their disagreement is retained rather than resolved by choosing the estimate nearest one.

The direct model has 140 deterministic recovery checks, twenty-eight joint-optimizer comparisons, 612 leave-one-state-out fits, thirty-six size/power scenarios and twenty-four trend-misspecification scenarios. The signed-score module adds eighty-four independent statsmodels nuisance-fit cross-checks and 140 deterministic conditional-mean recovery checks. These establish numerical properties, not exposure accuracy or causal validity. Survey-sampling uncertainty for these new mixture fits has not yet been propagated. Product type, unknown catch-up, outcome recording, geographic alignment and heterogeneous-effect transport remain sensitivity dimensions. No new independent participants are added.

S53.3 Japan's withdrawal trajectory

Japan's continuous coverage path can enter the same mixture mean. Four outcomes are profiled under constant, linear, quadratic and natural-spline baselines. The exposure series is a geographic proxy for the outcome population; nine cohorts cannot establish the counterfactual trajectory without restrictions.

All-ASD baseline	Fitted mixture R	Nominal profile set
Constant	0.111	lower search edge-0.723
Linear	0.829	lower search edge-4.288
Quadratic	2.075	lower search edge-10.396

All-ASD baseline	Fitted mixture R	Nominal profile set
Natural spline	1.738	lower search edge-10.387

The expanded grid runs from 0.000001 to 1,000. Every all-ASD lower endpoint reaches that search edge; it is not evidence of a positive lower bound. Endpoints away from the grid edge are refined by refitting the model at the crossing. Quasi-Poisson dispersion counts the estimated R as well as baseline parameters: n-rank(X)-1 residual degrees of freedom, or six for the linear and five for quadratic/spline baselines. The scaled profile-deviance chi-square approximation is nominal, not a finite-sample or serial-dependence guarantee. The constant baseline's apparent protective association is not preferred over the trend models.

Coverage's projection R-squared against the baseline rises from about 0.809 under the linear trend to 0.977 under quadratic/spline trends. This is loss of independent exposure information, not absence of a vaccination change. Twenty noiseless profile-containment checks pass; they are not power or point-identification tests. The native RR 0.981 per ten coverage points remains a different parameter from these mixture R estimates.

S53.4 Reproduction and synthesis

Run `processed/build_conversion_register.py`, `processed/pcv_exposure_aligned_profile.py`, `processed/pcv_signed_score_inversion.py`, and `processed/audit_japan_mixture_conversion.py` with the pinned environment and held inputs. The new conversion math tests include withdrawal reciprocity, zero contrasts, probability feasibility, open endpoints and weak first stages. Date-stamped receipts preserve input and code hashes. Four historical calibration/synthesis entry points require explicit archival opt-in; reproducing them is not endorsement of their exposure mapping.

A common-scale synthesis is permissible in principle when target outcomes, populations, time horizons and interventions are adequately aligned or modeled, and dependence is addressed. The present conditional alternatives do not establish one exchangeable vaccine effect. Priors, covariance corrections and repeated near-null specifications cannot substitute for the design-specific observation model. Neither a universal null probability nor a vaccine-class disability bound follows from these calculations.

S54. International original-observation likelihoods and hierarchical synthesis

S54.1 Observation blocks and exposure mapping

The international expansion combines eight vaccine-location contexts in seven original outcome blocks, with 1,944 outcome cells. U.S. PCV and RotaShield retain their joint likelihood on 353 state-cohort observations; the nine Japanese MMR cohorts form a disjoint outcome block. Five further panels contribute 1,582 cells. No historical causal-ready flag is used as an exclusion rule. Inclusion requires an observation model, while the strength of its causal interpretation is assessed separately.

Additional outcome system	Cells	Geographic units	Conditional outcome ratio, 95% CrI
Canada, newly identified autism	190	8 provinces	1.036 (0.362-2.055)
Sweden, F84 specialist care	756	21 counties	1.050 (0.774-1.585)
Germany, F84 inpatient discharges	36	10 states	1.034 (0.267-3.962)
England, school autism identification	88	National	0.946 (0.871-1.028)
Scotland, autism school support	512	32 local authorities	1.302 (0.339-4.520)

These are unpooled full-versus-no-exposure mixture parameters for recorded outcomes, with a Normal(0, log(2)-squared) prior on the log ratio. They are not automatically individual causal autism-risk ratios. The Canadian alternative prevalence likelihood gives 1.141 (0.384-1.951) on 192 cells; it replaces the newly identified-case likelihood, never multiplies it. PHAC uses newly identified cases to distinguish identification from biological onset (https://health-infobase.canada.ca/ccdss/Methods?wbdisable=true).

Canadian eligibility follows verified provincial program dates. Within the published ages 1-19 band, three mappings use uniform ages 1-19, uniform ages 2-8, and national pre-2010 case-age weights. Each is combined with 60%, 80% and 100% completed uptake. These nine equally weighted mapping scenarios represent uncertainty, not nine measured first stages. National preperiod age weights avoid estimating the weights from post-policy outcomes.

Sweden and Germany use ages 5-14, avoiding the assumption that completed infant coverage describes contemporaneous infants. Swedish coverage includes both published and imputed cohort values; HepB and rotavirus exposures enter as nuisance terms. Germany retains only fully supported RKI cohort windows. Its longer hospital archive therefore supplies only thirty-six matched older-age cells, and repeated discharges are not unique patients. The small counts limit the Gaussian log-rate approximation. England uses national age-specific totals once, excluding the nested local-authority totals; MenB and HepB eligibility enter as nuisance terms. Scottish exposure is school-age eligibility multiplied by average local coverage, not an observed historical receipt series.

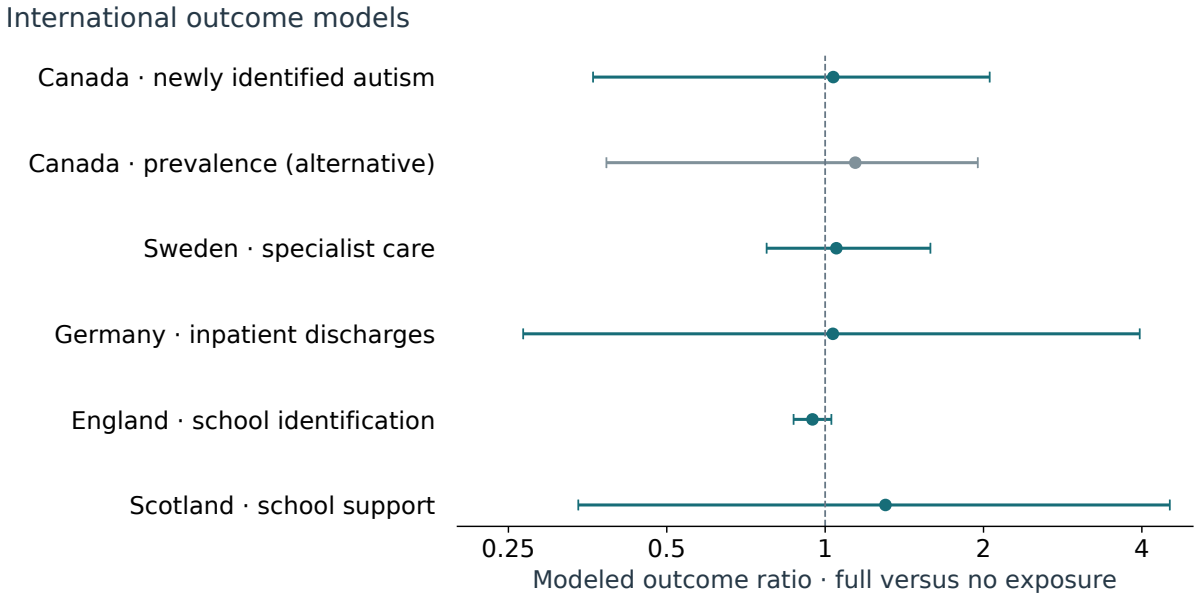


Figure S56. International original-panel mixture estimates. Points and bars show posterior medians and 95 percent credible intervals under the specified exposure mappings and log-ratio prior SD log(2), before the synthesis bias adjustments. Outcomes differ across rows. Canadian prevalence replaces, rather than supplements, the newly identified-case likelihood. These are recorded-outcome contrasts, not automatically individual causal autism-risk ratios.

S54.2 Working likelihood, heterogeneity and outcome bridges

Each additional panel uses a nonlinear mean proportional to 1 + p(exp(eta)-1), where p is its mapped exposure share and eta is its recorded-outcome log ratio. A proper Gaussian log-rate likelihood integrates nuisance coefficients and residual scale under a normal/inverse-gamma prior. Residual shape is 2.5 and prior root-mean-square scale is 0.25. The global intercept prior has relative scale 30; region, age and year intercepts have scale 20; region/age-specific linear time trends and natural-spline cohort terms have scale 2, all relative to residual standard deviation. These priors are regularization choices, not empirically validated counterfactuals.

Linear and four-degree-of-freedom natural-spline cohort baselines and temporal correlations 0, 0.6 and 0.85 receive equal initial weights. Their marginal likelihoods are averaged, not multiplied. The within-region residual kernel combines a temporal component of weight 0.8 with a repeated-cohort component of weight 0.2; the temporal component also permits cross-age correlation of 0.35. Population scaling is used when denominators are available. Alternative cohort/exposure mappings are likelihood alternatives on the same observations. Repeated diagnoses, vaccine-specific table copies and shifted-exposure copies are not additional outcome blocks.

The synthesis expresses each association parameter as a common mean plus product and context deviations and bias terms. Modern rotavirus and RotaShield are distinct product groups. Product heterogeneity SD takes 0.1, 0.35 or 0.7 and context heterogeneity SD takes 0.1 or 0.35, with equal prior mass on the six combinations; the likelihood updates these weights. The common mean has prior SD log(2). Context-bias SDs are 0.15 for U.S. contexts, 0.25 for Japan, 0.35 for Canada/England/Scotland and 0.70 for the two care-use bridges. Shared bias SD is 0.15, with additional shared terms of

0.20 for England/Scotland and 0.25 for the two care-use systems. Bias means are zero. These values are specified sensitivity assumptions, not measured bias corrections.

The exact likelihood curves enter numerical integration rather than a normal approximation to their posterior intervals. The existing joint U.S./Japan importance sample preserves PCV/RotaShield outcome dependence. Conditional Gaussian integration and grid convolution handle product borrowing and the foreign likelihoods. This construction follows the general principle of representing bias in mixed-source evidence synthesis, rather than treating all estimates as equally unbiased (Yao et al. 2024, doi:10.1111/jebm.12633); it is not their published estimator.

S54.3 Results, timing checks and limits

The reference exponentiated mean log effect is approximately 1.03 (95% credible interval 0.59-1.74). In the same hierarchy with only the U.S. and Japanese likelihoods, it is 1.02 (0.51-1.91). Adding the five international panels reduces the interval's width on the log scale by about 18%, conditional on this model. This quantifies added information about a hypermean, not an 18-percentage-point gain in causal certainty. The hypermean is not a population schedule RR or the probability that all context-specific effects equal zero.

Sensitivity analyses remove care-use bridges, double context-bias scales, vary the shared-bias and effect priors, substitute Canadian prevalence, change international trend/serial-dependence assumptions, and remove each outcome block in turn. An uncertain-bridge model gives each care-use system a 50% chance of following the autism effect with bias and a 50% chance of an unrelated association with its own Normal(0,1) distribution. Product-specific analyses avoid cross-product borrowing. The PCV-only hypermean is 1.15 (0.50-2.51); the modern-rotavirus-only hypermean is 1.02 (0.43-2.40). These are different targets from the shared hypermean. The new evidence does not establish tight universal individual-risk bounds.

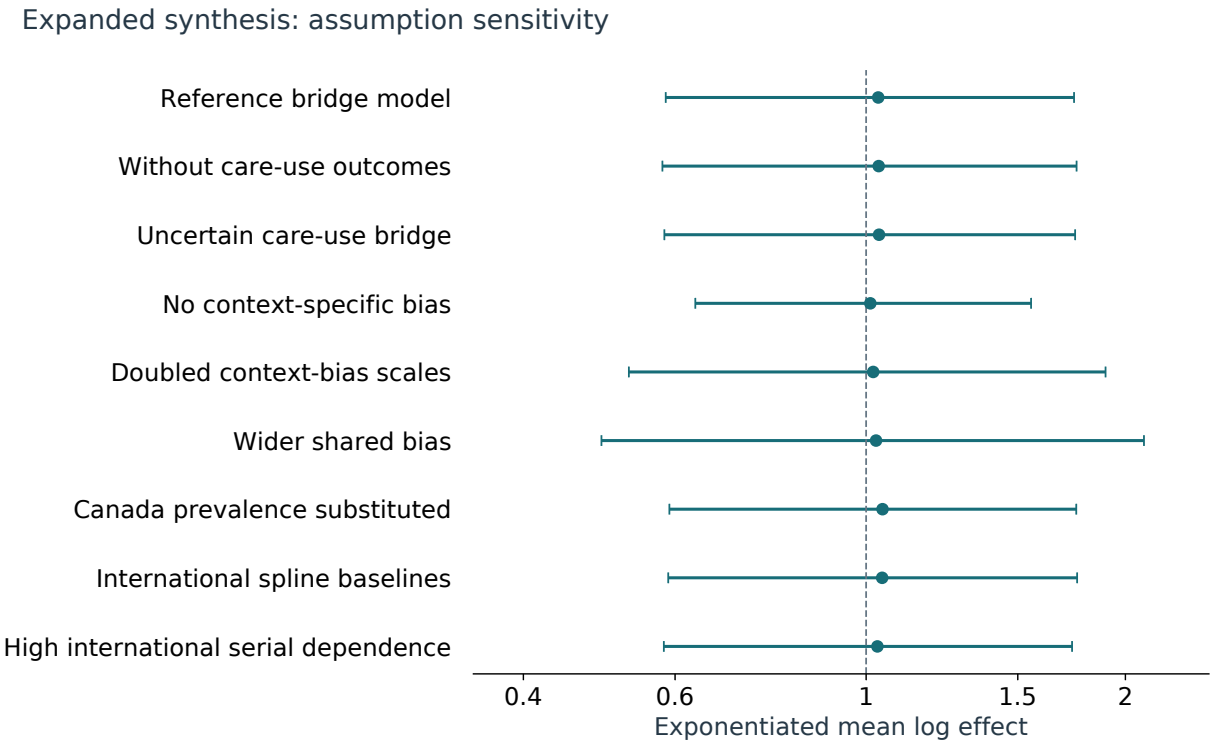


Figure S57. Expanded hierarchical synthesis sensitivity. Eight vaccine-location contexts contribute seven original outcome blocks. Product/context heterogeneity and shared bias are modeled. The plotted exponentiated hypermean is not a population schedule RR, a finite-context arithmetic average, or an all-effects-null probability. Care-use bridges and causal interpretations remain assumptions.

Three-year future exposure is compared with correctly timed exposure on identical outcome rows. Germany lacks sufficient matched support; its missing placebo is not a null result.

Paired timing test	Correctly timed ratio, 95% CrI	Future-exposure ratio, 95% CrI
Canada, 166 cells	0.793 (0.288-1.750)	1.600 (0.705-3.059)

Paired timing test	Correctly timed ratio, 95% CrI	Future-exposure ratio, 95% CrI
Sweden, 630 cells	1.004 (0.734-1.374)	0.911 (0.728-1.236)
England, 64 cells	1.007 (0.903-1.125)	1.016 (0.874-1.120)
Scotland, 416 cells	1.175 (0.294-4.407)	1.112 (0.298-3.853)

These intervals are broad and do not establish timing specificity. England's full-window precision does not carry over unchanged to the shorter matched comparison. Future exposure can proxy persistent regional characteristics; it cannot cause an earlier outcome. The paired results are not independent of the primary fits and do not enter the hierarchy again.

S54.4 Reproduction and remaining integration

A common-main-effect sensitivity sets product and context heterogeneity to zero while retaining the stated exposure mappings, context/shared bias uncertainty and separate U.S. excess interaction. It gives 1.082 (0.731-1.613). Thus the analysis also permits homogeneous main effects; its broad reference interval is not assessed only under models requiring nonzero heterogeneity.

Run `processed/international_likelihoods.py`, `processed/international_hierarchy.py`, and their report/validation scripts. `project_results/international/inputs.json` retains original modeled observations, exposure scenarios and source hashes. Analytic Gaussian benchmarks, no-data recovery, likelihood-constant invariance, positive-definite kernels, unique observation keys, denser likelihood grids and independent numerical scrambles test implementation. These checks do not certify causal assumptions or complete the project-wide evidence synthesis.

This eight-context comparison represents underlying panels from five crosswalk files. The eleven-context model in S55 represents seven crosswalk source files; the count-data extension in S56 partly represents nine of the sixty-six files, leaving fifty-seven without implementations in that hierarchy. England's fixed-age and Quebec's age-specific panels are additional to those source-file mappings. An unfinished implementation is not evidence that a source is unusable. This autism-oriented hierarchy does not synthesize all chronic disease or disability outcomes.

S55. Expanded observation systems, overlap replacement and outcome-specific integration

S55.1 Additional autism panels and shared dependence

The expanded model includes eleven vaccine-location contexts, ten original outcome blocks and 2,260 outcome cells. Northern Ireland contributes 155 grade-year observations and Hong Kong contributes 60 grade-year observations. Quebec contributes 125 observations across five nonoverlapping age bands and twenty-five years. Quebec's 24 observations in the Canadian newly identified-case block are removed; the remaining seven provinces contribute 166 cells. Thus no Quebec surveillance outcome is multiplied by its age-specific version. The shared U.S. PCV/RotaShield likelihood and all other S54 outcome blocks remain.

Additional autism panel	Cells	Conditional outcome ratio, 95% CrI
Canada, seven provinces, newly identified cases	166	0.892 (0.275-1.776)
Northern Ireland, school autism	155	1.012 (0.878-1.155)
Quebec, age-specific autism prevalence	125	1.057 (0.703-1.508)
Hong Kong, school autism	60	0.892 (0.657-1.165)

These are unpooled recorded-outcome mixture ratios with the log-ratio prior SD log(2), before the synthesis bias terms. Northern Ireland combines eligibility with uptake scenarios of 80%, 90% and 100%. Age-specific source-era nuisance terms allow the 2019 change from the SEN to medical register to differ across ages; MenB and HepB eligibility also enter as nuisance terms. Published and digitized rates have limited precision. The source authority warns that the two reporting eras are not directly comparable (Northern Ireland Department of Health statistical dashboard, https://datavis.nisra.gov.uk/health/iad-statistical-dashboard.html).

Quebec's rotavirus programme began in November 2011 (Comeau et al., PMID:26795368). Eligibility is averaged within each age band with alternative birth cutoffs 2011.25, 2011.75 and 2012.25, and uptake 60%, 80% and 100%. These are alignment and receipt scenarios, not measured first stages. Age-specific 2016 billing-level and slope changes, plus pandemic indicators, are nuisance terms. Crude published prevalence is rounded and concerns recorded diagnoses, not biological onset.

Hong Kong uses published PCV coverage for the 2009-2012 cohorts and three earlier-exposure scenarios: routine-only, 20% prior private receipt, and additional catch-up coverage of 50% and 90% for the 2007 and 2008 cohorts. These historical values are sensitivity assumptions, not verified uptake. The numerator covers public-sector ordinary-school SEN while the denominator covers all local ordinary schools. Grade-based birth cohorts and nonlinear cohort trends remain uncertain. The rate likelihood does not repair that scope mismatch.

The hierarchy extends the UK shared-bias term to Northern Ireland. Quebec and the other Canadian provinces share a further Normal(0, 0.20-squared) log-bias term. Hong Kong shares the PCV product effect with the U.S., Sweden and Germany but has a separate context-bias SD of 0.70. Northern Ireland and Quebec use context-bias SD 0.35. Full foreign likelihood curves, rather than posterior normal approximations, enter the existing joint-core integration.

S55.2 Synthesis and timing sensitivity

The denser reference calculation gives an exponentiated mean log effect of 1.018 (0.600-1.658). Flexible international cohort trends give 1.030 (0.604-1.675); high international serial dependence gives 1.014 (0.595-1.651); doubled context-bias scales give 1.002 (0.536-1.818). Removing Hong Kong gives 1.028 (0.600-1.682), removing Northern Ireland gives 1.020 (0.591-1.679), and replacing Quebec's age bands with the original CCDSS contribution gives 1.027 (0.596-1.688). These are conditional hypermeans, not individual schedule RRs or probabilities that every effect equals zero.

The 14 reported synthesis summaries include prior and bias sensitivity. Matched three-year future-exposure tests reuse identical outcome observations within each comparison. They remain diagnostics, never additional likelihood contributions. Full native intervals, alternative trend and serial settings, and all paired timing results are in the linked integration report. The shortened matched windows do not retain the precision of the complete panels.

S55.3 Other disability and utilization outcomes

The same observation-likelihood framework is fitted separately to eight additional Hong Kong categories and five Danish somatic hospital-use outcomes. The Danish contrast is earlier second-MMR eligibility at age four rather than twelve, not MMR introduction or withdrawal. Ages two and three form never-second-dose-exposed comparisons; ages six, eight and ten become eligible under the earlier schedule. Three historical age alignments and the same trend/serial alternatives are fitted. Published rates are integer-rounded and actual receipt is not reconstructed.

Additional outcome	Conditional ratio, 95% CrI
Hong Kong, ADHD	1.096 (0.843-1.285)
Hong Kong, hearing impairment	0.986 (0.716-1.323)
Hong Kong, intellectual disability	1.460 (1.178-1.917)
Hong Kong, mental illness	0.517 (0.254-1.585)
Hong Kong, physical disability	1.783 (1.226-2.903)
Hong Kong, specific learning difficulty	0.972 (0.602-1.620)
Hong Kong, speech/language impairment	0.567 (0.277-1.046)
Hong Kong, visual impairment	2.009 (0.560-6.058)
Denmark, outpatients	0.968 (0.866-1.077)
Denmark, outpatient treatments	0.971 (0.861-1.089)
Denmark, admissions	1.002 (0.894-1.122)
Denmark, inpatients	1.001 (0.903-1.107)
Denmark, bed days	1.022 (0.905-1.147)

The intellectual- and physical-disability reference intervals exclude one, and their positive signs persist across the fitted trend and serial variants. They must not be characterized as null results. On the shortened matched timing samples, intellectual disability is 1.170 (0.382-2.359) for correctly timed exposure and 0.882 (0.327-2.011) for future exposure; physical disability is 1.874 (0.877-3.869) and 0.639 (0.337-2.229), respectively. These comparisons are much less precise. The intervals are marginal and do not adjust for the outcome search. Small counts, only five Hong Kong school years (two for mental illness), possible classification shifts, unmeasured earlier vaccination and denominator mismatch preclude treating the associations as identified vaccine harms. The positive results also prevent an assertion of universal observed nulls for disability.

Each outcome retains its own estimand. Hospital-use rates can count repeated care and are not disability incidence. Hong Kong categories overlap in observation systems and potentially in pupils. Neither these outcomes nor the five Danish endpoints are multiplied into the autism hypermean or pooled as a single disability effect.

S55.4 Whole-project integration audit and reproduction

The audit accounts for all 531 current result-source entries (50 of which hold only derived posterior summaries and are classified as not additional evidence) and every source in the 66-file international crosswalk. It is a result-file disposition map, not certification that all acquired raw data or all causal estimands have been integrated. Nine crosswalk files have partly refitted underlying panels when S56 is included; fifty-seven remained without implementations in that hierarchy. S59 now gives every file a disposition. Other exposures and outcomes within represented files also remain. The Norwegian primary-care age-gradient models in S58 are a separate outcome family and are not inserted into an autism hypermean.

Published individual-study evidence already has family-specific meta-analyses and overlap sensitivities; its separate role is not an accidental omission. NSCH supplies individual outcomes but state/cohort vaccination exposure, so it does not become an individual vaccination comparison. EURODIAB, Finnish diabetes, YRBS and WIC remain outcome-specific analyses. DNPAO and DASH presentations of YRBS overlap and cannot be added as independent samples. Interactions over the same outcome cells require a joint multiexposure model, not an additional independent estimate.

Some remaining integrations require data corrections or new observation models: Taiwan and certain Korean outcomes measure disability composition, not population autism rates; Czech absent count strata lack a verified publication rule (the S56 sensitivity retains that uncertainty); Brazilian administrative vaccine coverage can exceed 100%; and Norwegian/Finnish support or primary-care outcomes differ from autism. These issues are documented separately from ordinary unfinished implementations. They are not grounds to discard all information or to declare a null.

`processed/integration_expansion.py` constructs the held-observation inputs and likelihoods. `processed/integration_hierarchy.py` performs the expanded synthesis. `project_results/integration_audit/code/reproduce_integration.py` replays frozen inputs and the power-16 hierarchy. The reference uses power 18; seven unit tests include an analytic 12-dimensional Gaussian benchmark for the product, context, UK, Canadian and care-system dependence structure. Source hashes, overlap removal, unique original keys, exposure bounds, positive-definite kernels, grid convergence, independent scrambles and local report links are checked separately. Numerical reproducibility is not causal certification, and this exploratory model has not received independent methods review.

S56. Czech count likelihoods and Finnish pre-primary support

S56.1 Czech outcomes, population denominators and missing strata

The NZIP outcome is the annual number of unique children receiving recorded psychiatric or psychological care with an F840 diagnosis, not new diagnoses or all autism in the population. The panel has 1,078 cells from 77 districts, outcome years 2016-2024 and ages 4-6 or 7-10. Its exposure is MMR recorded by age two, not dose-two coverage or lifetime vaccination. Ten annual CSO age/sex files, 2015-2024, supply matching resident-population offsets. Prague is represented by its region total, without adding the overlapping district total. Birth counts and averages of adjacent year-end populations provide separate offset sensitivities. The actual-population to birth-count ratio ranges from 0.86 to 1.46, with median 1.02.

Sources: NZIP autism-care data and documentation, NZIP vaccination documentation, and CSO annual age/sex population files. Receipt by age two ranges from 65.1% to 90.6% across retained cells. ISIN vaccination reporting augments the reimbursed-care register from 2023, so reporting changes may affect the recorded first stage. A pre-2023 outcome-window sensitivity is retained, although it cannot undo retrospective reporting changes to exposure cohorts.

Five combined cells are zero after filling absent strata, but 194 sex-specific strata are absent. The publication does not explicitly establish an absence-means-zero rule. The primary calculation assigns zero; separate sensitivities assign one or two to every absent sex stratum. Those calculations do not validate the publication rule and are not averaged as if they were observations from different studies. Male/female totals are combined once. F840, the all-F84 total and diagnostic subtypes cannot be multiplied as independent outcomes.

The vaccination workbook contains fourteen birth-cohort coverage sheets, five annual-count sheets and three calendar-year coverage sheets. Only the first group permits the cohort adapter. MMR-by-two and the hexavalent/childhood-PCV sheets used for the existing interaction models pass this header contract. Annual counts are not percentages, and a calendar reporting year is not a birth year. The typed source audit retains all sheets while preventing those unit errors from entering cohort models.

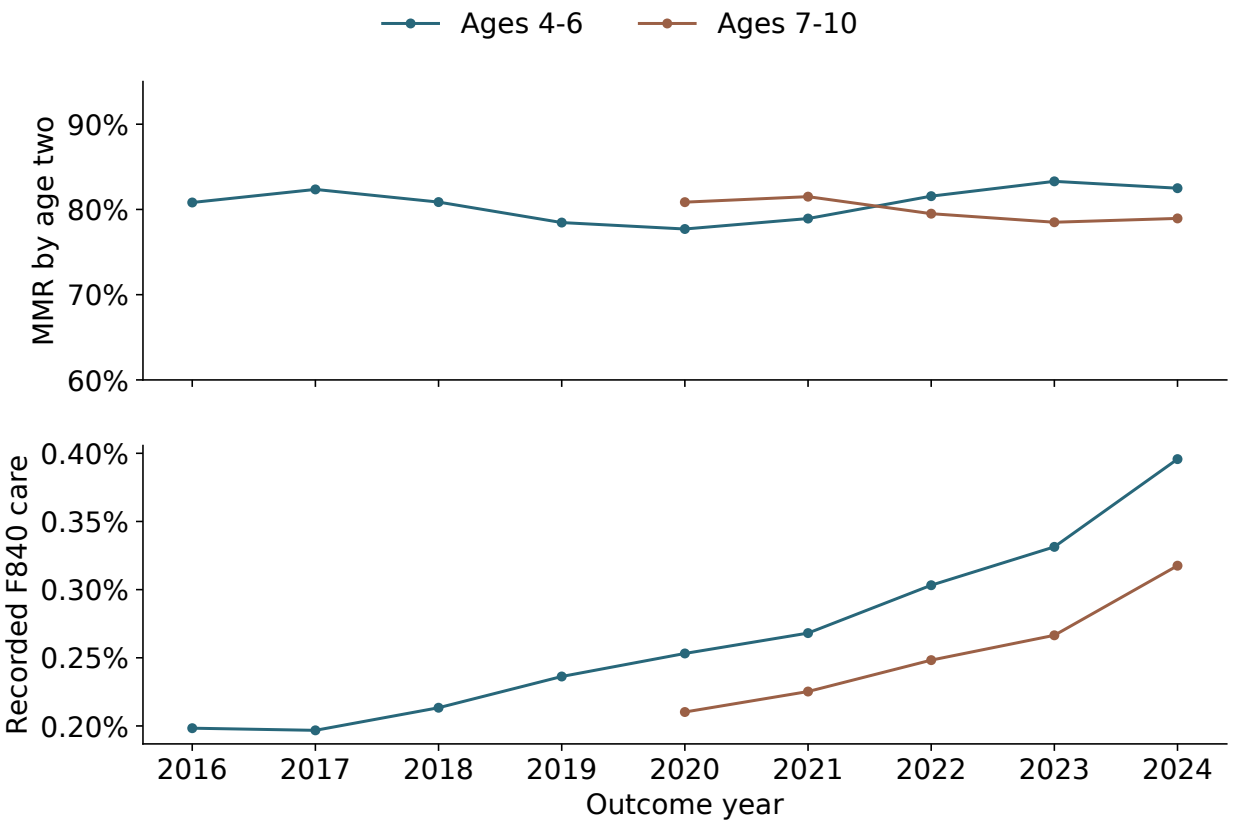


Figure S58. Czech cohort-aligned recorded MMR receipt by age two and F840 care rates. The age-band trajectories use age-matched resident populations and the primary absent-as-zero count assignment. These are descriptive patterns, not a policy discontinuity.

S56.2 Working model and sensitivity

The expected count is the resident-population offset times a latent baseline care rate times the mixture factor 1 + p(R - 1), where p is cohort-matched receipt by age two and R is a conditional exposed-to-unexposed care-risk ratio. Transporting R to an individual autism effect additionally requires exchangeability, valid receipt mapping, a common baseline-risk mixture and ascertainment assumptions. Later receipt among children not vaccinated by two, migration, diagnostic access and shared healthcare reporting can violate that interpretation.

The baseline log rate has proper Gaussian terms: a global intercept (SD 3), district-by-age intercepts (SD 2), calendar-by-age effects (SD 1), district slopes (SD 1 on centered time scaled by half the observed range), district-year AR(1) deviations (stationary SD 0.5), shared district-birth-cohort deviations (SD 0.5 with equal weights over

contributing birth years) and observation-level deviations (SD 0.2). The spline variant adds a centered four-degree-of-freedom natural cohort spline with coefficient SD 0.7. These are regularization assumptions, not estimated precision parameters. The reference averages six models: linear/spline trends crossed with AR correlations 0, 0.6 and 0.85, using equal prior model weights and integrated-likelihood weighting.

Gaussian nuisance terms are integrated approximately by Laplace's method. The ratio likelihood is evaluated on 121 log-ratio points from -6 to 6. The reference log-ratio prior has mean zero and SD log(2); SD 1.2 and a mean of log(1.25) test effect-prior sensitivity. Other specifications change absent-stratum assignment, the offset, age band, outcome window, district-trend prior or serial variability. A three-year shifted exposure window is a timing diagnostic, not a guaranteed untreated control: the ages-7-10 windows overlap one birth year. The reference ratio is 1.035 (0.322-3.660), showing little information about small effects. The finer-grid representative model has BF01 about 1.11 and prior-to-posterior information gain about 0.013 nats; neither is strong evidence for an exact null.

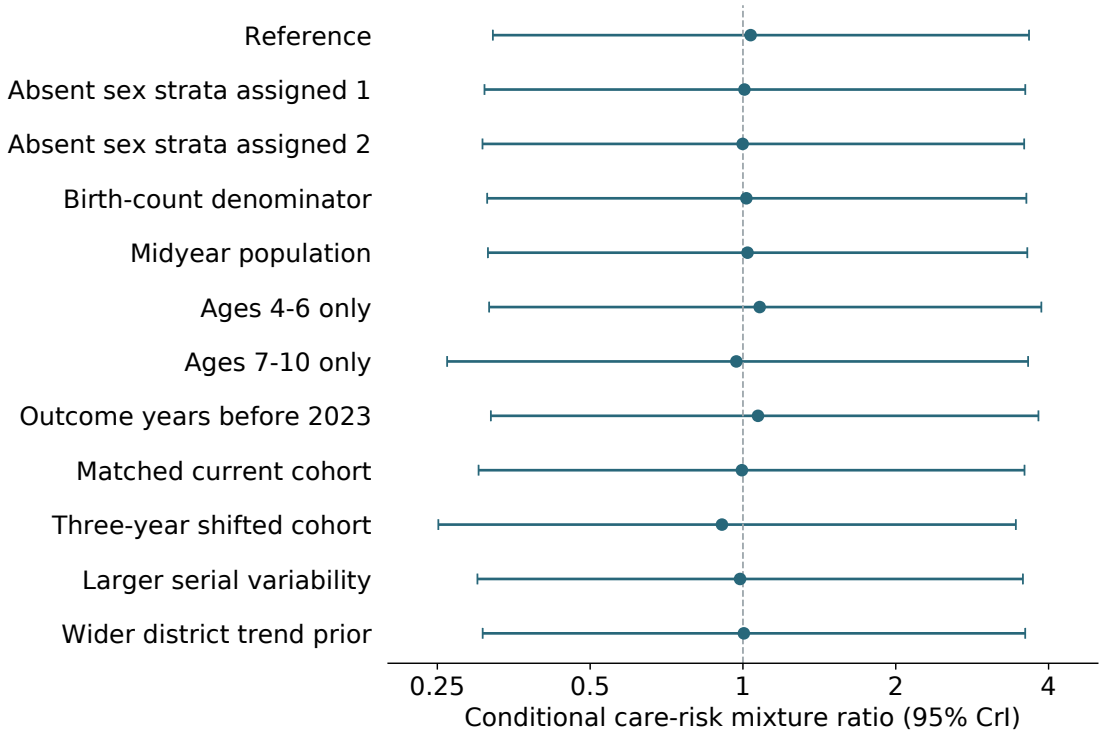


Figure S59. Czech count-model sensitivity estimates. Intervals are 95 percent conditional credible intervals; the shifted-cohort row is a timing diagnostic, not an independent effect. The broad intervals and prior dependence limit individual-risk inference.

S56.3 Integration without multiplying reused outcomes

Czechia adds one joint count likelihood to the MMR product component shared with Japan. A context-bias SD of 0.70 on the log scale acknowledges its care-use and receipt-by-two bridge. Doubling that bias SD and allowing a 50% unlinked-context mixture test the bridge, without claiming to encompass every plausible bias. The U.S. PCV/RotaShield joint likelihood and the UK, Canadian and care-system shared terms remain intact. The two Czech age bands are not independent replications.

The resulting twelve-context, eleven-block model has 3,338 original cells. Its power-18 shared-mean ratio is 1.023 (0.610-1.661), versus 1.018 (0.600-1.658) without Czechia. The small change in interval width reflects modest information gain under the model, not proof of a tightly bounded individual effect. The reference posterior means a common hypermean over heterogeneous contexts; it is neither a schedule-wide population RR nor the probability that every product effect is zero. Finnish support outcomes do not enter this autism likelihood.

S56.4 A distinct Finnish observation setting

StatFin table 147s covers pre-primary education in early-childhood education and care, separate from the existing school-based table 137q. It supplies 191 municipalities in 2022-2024 and 134,984 pupil-years before exposure filters. Municipality sums reconcile with each national total for pupils, special support and intensified support in each year. Special support is an educational provision, not an autism diagnosis or a comprehensive disability measure. The MMR1 special-support analysis retains 514 cells, 133,762 pupil-years and 5,769 support pupil-years.

The table excludes Aland. Its regional units describe educational location under the latest-year municipal classification, not necessarily pupil residence. A separate JSON-stat reacquisition matches all 5,184 published values, including the five intensified-support denominator conflicts. Agreement between two representations verifies parsing of the same source; it is not independent outcome replication.

THL exposure metadata distinguish timing windows. MMR1, rotavirus, PCV, DTaP-IPV-Hib and varicella measures follow receipt through age three. MMR2 follows ages four to eight and the DTaP-IPV booster follows ages two to eight: these extend beyond the age-six outcome. Those two associations remain visible but are not antecedent-exposure tests. All measures use municipality at the atlas update, which need not be the municipality at vaccination or preschool attendance. Continuous-residence eligibility, reporting completeness and suppressed small cohorts add measurement and selection limits.

Binomial-logit models retain actual numerator/denominator counts, municipality intercepts, calendar effects, municipality trends and serial terms with proper Gaussian nuisance priors. Coefficients are ecological odds ratios per ten percentage points of recorded coverage, not individual vaccination ORs. Specifications vary serial correlation, reporting completeness, matched/future cohorts and adjustment for three other coverage measures. The latter is a correlated-coverage sensitivity, not a test identifying biological synergy. Eight reference exposure summaries and all specification/prior variants remain in the companion report; two references have the post-outcome timing limit. Posterior quadrature uses a 2,401-point cubic interpolation of the 161-point fitted likelihood.

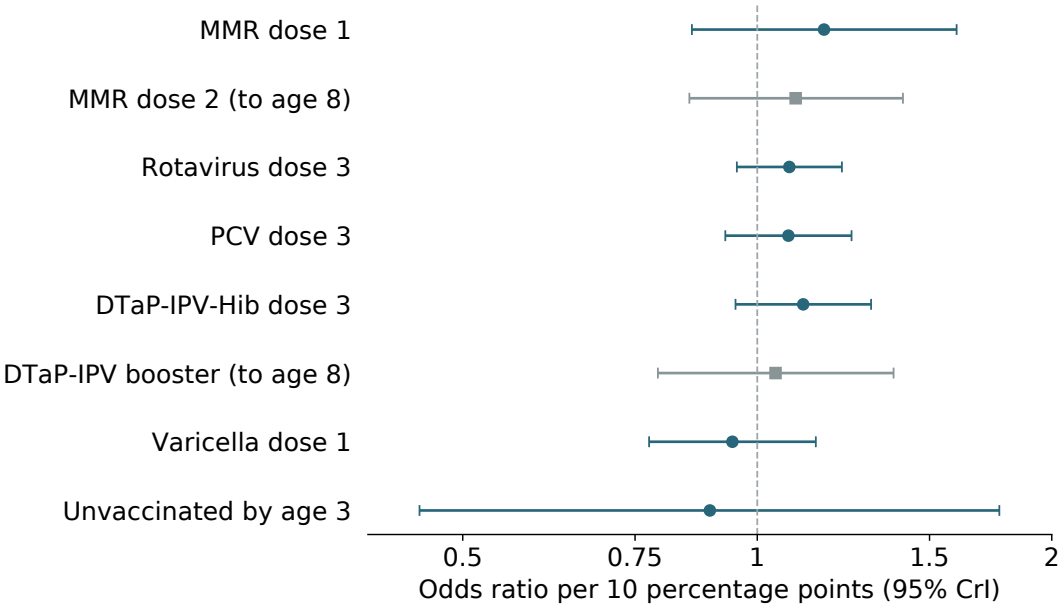


Figure S60. Finnish ECEC special-support gradients. Ratios are ecological odds ratios per ten percentage points of coverage, not individual vaccination ORs. Gray squares have exposure follow-up extending beyond the outcome age. The same pupils contribute to multiple models, so rows are not independent replications.

Five intensified-support counts exceed their reported denominators: Marttila 2022 (8/1), Kangasala 2022 (67/53), Tampere 2022 (226/55), Tampere 2023 (228/72) and Sodankyla 2023 (15/12). No binomial effect is reported for this outcome. Counts are neither clipped nor selectively dropped. A source-definition reconciliation is needed before treating these counts as subsets of the reported pupils. Even the valid special-support counts have only three outcome years and do not identify an exogenous vaccine-policy contrast.

S56.5 Numerical checks and reproduction scope

Unit tests compare one-dimensional Laplace integrals with numerical integration, recover known injected effects, verify AR covariance and birth-cohort sharing, and recover the eleven-context hierarchy when the added Czech likelihood is flat. Representative twice-dense Czech and Finnish fitting grids leave the reported posterior quantiles within the specified numerical tolerances. These checks do not establish accuracy of every high-dimensional Laplace approximation or causal validity. The portable `count_extension/code/reproduce_count_extension.py` rebuilds the held-input native likelihoods in a fresh cache; `--hierarchy` also reproduces the power-16 synthesis. Exact source hashes, acquisition scripts, missingness decisions and isolated replay receipts accompany the results. Raw-data reacquisition, numerical reproduction and causal identification are separate claims.

S57. Scottish school-support counts and outcome-specific diagnostics

S57.1 Source definitions and confidential counts

The Scottish Government pupil census supplies sixteen annual workbooks, 2010-2025, covering primary, secondary and special schools. A source audit checks 9,216 authority-year-sector-outcome entries for autism, hearing, visual, physical/motor, speech/language and learning-disability support. These are reasons for educational support; pupils can have multiple reasons, and a recorded support category is not incident clinical disease. All six outcomes are required by the parser. The primary-school hearing header in 2011-2017 is spelled "Hearing impaiment" and is explicitly recognized.

The workbooks contain 252 confidential sector counts. Their notes distinguish small-cell suppression from additional suppressions used to prevent disclosure. Consequently, four is not a verified upper bound for every suppressed count. The count-model reference retains authority-year outcome cells with no confidential sector component. Separate assignments of 0, 2, 4 and 25 per confidential component examine sensitivity; none is a disclosure bound. The propensity-model comparison assigns 2.5, a midpoint scenario rather than an observed count. Source not-applicable symbols (# or z) are assigned zero in the no-sector setting; the raw tokens remain available. Selection into fully reported cells can be nonrandom and is not solved by exclusion.

The six fully reported samples contain respectively 490, 420, 451, 491, 489 and 498 cells, out of 512 per outcome. The underlying pupils, authorities and reporting systems overlap across outcomes. National totals, school sectors and support reasons are not independent studies.

S57.2 Exposure and count likelihood

Exposure is a proxy school-age eligibility share multiplied by mean 2020-2025 local rotavirus uptake from Public Health Scotland. It is not observed historical receipt in the first eligible cohorts. The eligibility ramp begins with school entry in 2018, five years after introduction, and assumes equal weight over ages 5-18. Actual special-school ages, migration and within-authority uptake can differ. Cohort random effects use the same broad age weights as a working dependence model, not an exact record of pupil birth cohorts.

Conditional on proper Gaussian nuisance effects, counts follow a Poisson distribution with pupil-population offsets and the risk-mixture multiplier 1 + p(exp(eta) - 1). The nuisance structure includes authority intercepts, year effects, authority-specific linear trends, authority-year serial variation, overlapping birth-cohort effects and observation-level variation. Laplace integration approximates the nuisance integral. The reference averages six likelihoods: serial correlations 0, 0.6 and 0.85, each with and without an additional common natural cohort spline. In this single broad-age group the common cohort spline is collinear with calendar-year effects: it changes their prior, not the mean-design span. A separate sensitivity adds authority-specific quadratic and cubic trends, residualized on local intercept and linear time and assigned Gaussian coefficient SD 0.5 on time scaled to [-1,1]. Those trends remain regularized, not unrestricted counterfactual paths.

The main grid contains 78 actual count fits and 102 likelihood settings, each examined under centered log-effect priors with SD log(2) or 1.2 and a harm-centered prior with mean log(1.25) and SD log(2): 306 summaries. Six local-nonlinear fits add eighteen summaries. Lower proxy exposure, confidential-count assignments, serial structure, trend structure and matched three-year-ahead exposure are reported together. Future exposure remains correlated with the observed ramp, so this is a timing diagnostic rather than an independent placebo experiment.

S57.3 Estimates and event paths

The reference conditional support-count mixture ratios are 1.55 (95% credible interval 0.38-5.33) for autism, 1.17 (0.31-3.91) for hearing, 1.15 (0.30-3.85) for visual, 1.05 (0.28-3.46) for physical/motor, 1.13 (0.30-3.78) for speech/language and 0.91 (0.25-2.95) for learning-disability support. All intervals are broad. Point-null Bayes factors against the reference continuous alternative range from 0.80 to 1.10, adding little information. These are neither tightly bounded individual effects nor demonstrated vaccine harms. The local-nonlinear autism sensitivity is 1.72 (0.43-5.61); its imprecision reinforces that conclusion.

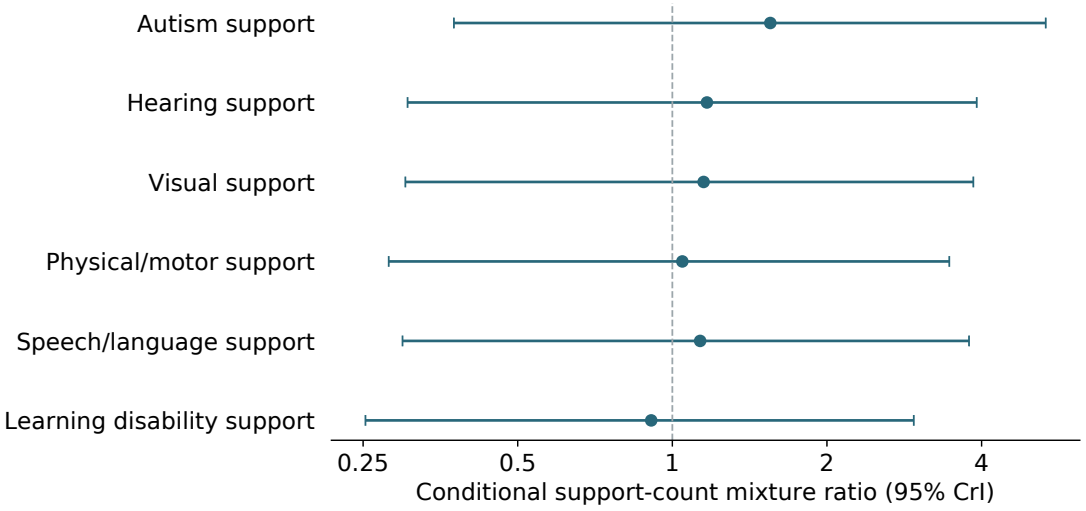


Figure S61. Scottish school-support count models. Posterior medians and 95 percent credible intervals use fully reported cells and the reference effect prior. Exposure is a school-age eligibility proxy multiplied by later area uptake. These six outcomes share pupils and administrative systems; they are separate, conditional recorded-support contrasts, not six independent causal replications.

Each support outcome receives its own event-study lead test. The seven-lead clustered Wald chi-square p-values, using midpoint-assigned versus fully reported cells, are: autism 0.137 versus 0.133; hearing 0.009 versus 0.042; visual 0.129 versus 0.103; physical/motor 0.684 versus 0.726; speech/language 0.253 versus 0.021; and learning disability 0.673 versus 0.676. These are asymptotic diagnostics with 32 authority clusters, not exact small-sample validity gates. Hearing rejects in both samples, and speech/language rejects in the fully reported sample. Autism's non-rejection cannot be assigned to other outcomes. The event-study coefficients below use a different scale from the mixture model: log rate ratios per standard deviation of persistent uptake, relative to 2017, with marginal t31 intervals.

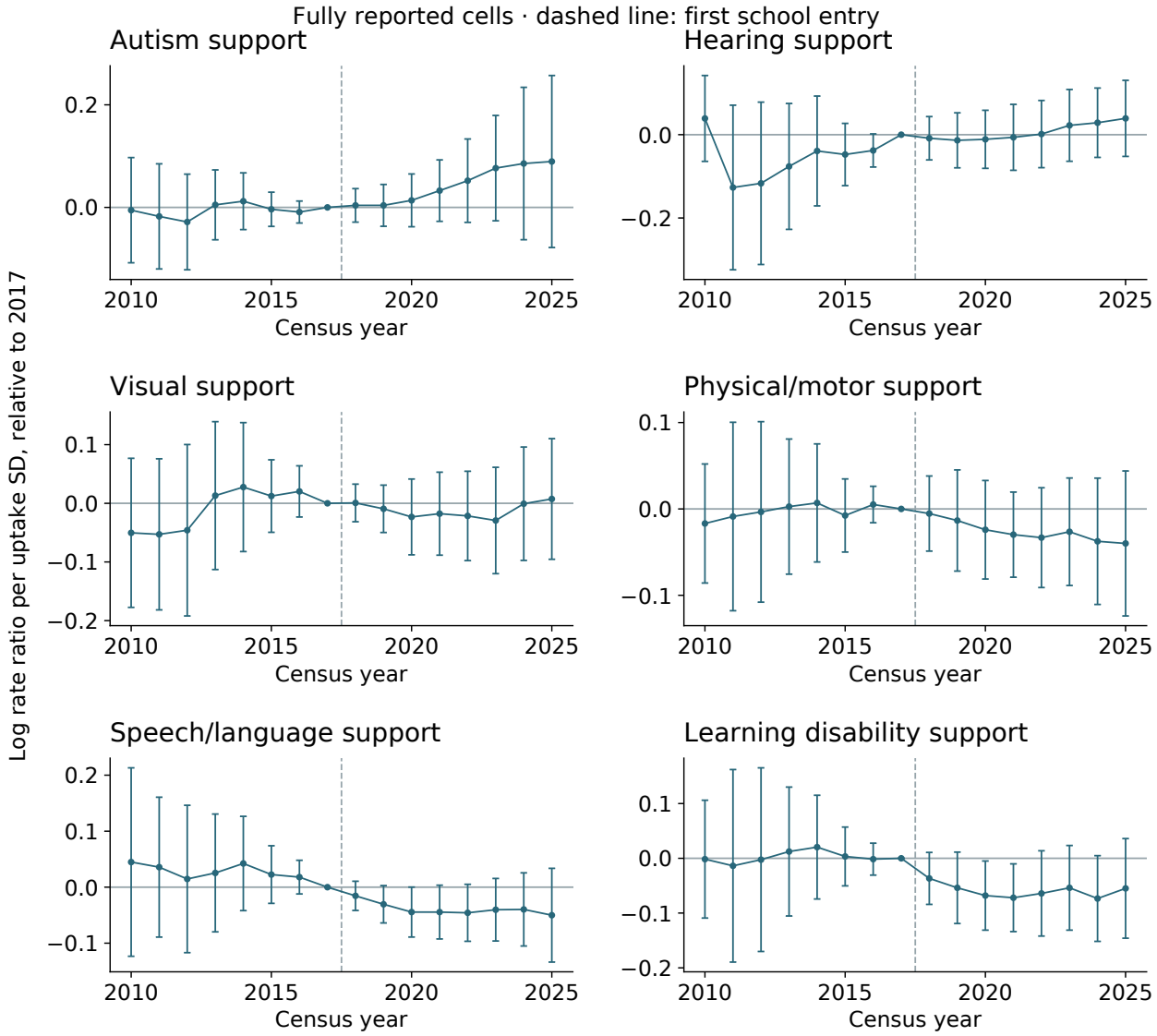


Figure S62. Scottish outcome-specific event studies. Each panel uses its own fully reported sample, with 2017 as reference. The dashed line marks first school entry of eligible cohorts, not the vaccine-introduction date. Marginal 95 percent intervals use local-authority clustered covariance and t31 critical values. Ratios are per standard deviation of persistent area uptake. Hearing and speech/language pretrends reject in these samples; see S57 for the count-assignment comparison.

S57.4 Overlap-safe synthesis and numerical checks

Replacing the 512-cell Gaussian Scottish autism block in S56 with the 490-cell fully reported count likelihood retains twelve vaccine-location contexts and eleven outcome blocks, now with 3,316 original cells. The shared UK bias term is retained. No non-autism support outcome enters the autism synthesis. The conditional shared-mean ratio is 1.028 (0.613-1.670), compared with 1.023 (0.610-1.661) under the imputed-count specification. These are alternative observation assumptions, not likelihoods to multiply. They do not materially narrow the cumulative interval or establish a small upper bound for every vaccine or outcome.

The replacement builder recovers the original international likelihood when supplied the original Scottish curve, across two context-variation and two bias settings. Power-16 and power-18 integration and two independent quasi-Monte Carlo scrambles give stable reported summaries. Three representative count likelihoods were also refitted on a 241-point grid instead of 121 points; the largest absolute change in a reported ratio quantile is approximately 0.012. Six independently restarted null modes from the packaged code agree within 0.000011 log-likelihood units at the specified Newton stopping tolerance. Source-header unit tests require all six outcome columns and recover the misspelled hearing

header. These checks address code, source and numerical behavior, not causal identification or exactness of all high-dimensional Laplace approximations.

The companion `scotland_support` directory supplies source tokens and workbook hashes, complete native and event-study results, corrected display tables, SVG figures, frozen observations, code and a fresh-cache reproduction command. Its source audit retains the original tables separately. Historical result rows are not counted again as new evidence.

S58. Norwegian childhood-versus-adult primary-care gradients

S58.1 Full-age acquisition and outcome meaning

The FHI KPR public table 370 supplies annual unique-person primary-care counts and unstandardized rates for eleven ICPC condition groups. A complete age-dimension request yields 984,720 source rows, including 357 municipalities, three sex categories and ten published age bands in 2017-2024. The 196,944 rows shared with the two-age extract match exactly for counts, rates and disclosure flags. The expansion therefore adds 787,776 published source rows, not that many independent people or treatment assignments.

The model uses six disjoint age bands rather than counting nested totals as additional independent observations. Its frozen inputs contain 414,273 published municipality-year-sex-age-condition cells and 14,784 age-two coverage cells from SYSVAK table 648. Psychiatric symptoms, diagnostic groups, musculoskeletal conditions, cardiovascular diagnoses and injuries are all claim-relevant outcomes; none is stipulated to be a negative control. They overlap within patients and do not measure autism-specific incidence or comprehensive disability. Source notes identify effects of diagnostic practice, private-care exclusions, geography changes, small-cell and secondary suppression, and suppression of entire sparse series.

Unstandardized RATE and TELLER imply a common population denominator across condition groups in each geography-year-age-sex cell, with maximum relative spread below 0.00000000000001. Some denominators are half-integers. They are retained as population weights, not rounded into binomial trials. Confidential observations are not assigned zero.

S58.2 Age contrasts, specifications and inference

The reference dependent variable is the difference between log primary-care rates at ages 0-14 and ages 45-64 within the same municipality and calendar year. It is regressed on age-two area vaccine coverage, scaled per ten percentage points, with municipality and year effects and harmonic age-group population weights. Exponentiating the coefficient gives a ratio of child/adult care-rate ratios, not a vaccinated/unvaccinated risk ratio. Age differencing removes only influences entering the two age groups in the same way. It cannot eliminate age-specific recording incentives, service changes or secular trends. Comparison ages are not presumed unvaccinated, and coverage is not matched receipt for the full 0-14 age band.

There are 77 primary vaccine-outcome comparisons. The full grid has 4,851 settings across five comparison age groups, contemporaneous/future/one- and two-year-lag exposure, local intercepts or linear/quadratic trends, sex-specific models and unweighted sensitivity. A model requires at least four plus the local polynomial degree observed years per retained municipality and at least ten municipalities. Thus 4,686 settings are estimable; 165 remain explicitly unestimated rather than coded as nulls. Lag and future-exposure settings can change the sample and are timing diagnostics, not randomization tests.

Weighted alternating projections remove the municipality-polynomial and year nuisance spaces. Dense explicit-regression tests recover the same coefficients and municipality-clustered variances. Both municipality and county clustered inference is retained; the primary display uses county clustering, t critical values and a CR1 finite-sample multiplier based on an upper bound on nuisance rank. There are fifteen county clusters in the primary comparisons. The model accounts for dependence within these clusters but does not guarantee elimination of all cross-county shocks or correct counterfactual trends.

S58.3 Results, precision and scope

Primary point estimates range from 0.802 to 1.447 on the native age-gradient scale. Eight of 77 primary p-values are below 0.05 before adjustment; none survives Benjamini-Hochberg correction across that family (minimum q=0.312). Another 151 nominal associations appear in the correlated sensitivity results and remain in the register. These are not a universal replication of exact nulls. Neither an adverse-sign coefficient nor non-rejection identifies a biological vaccine effect.

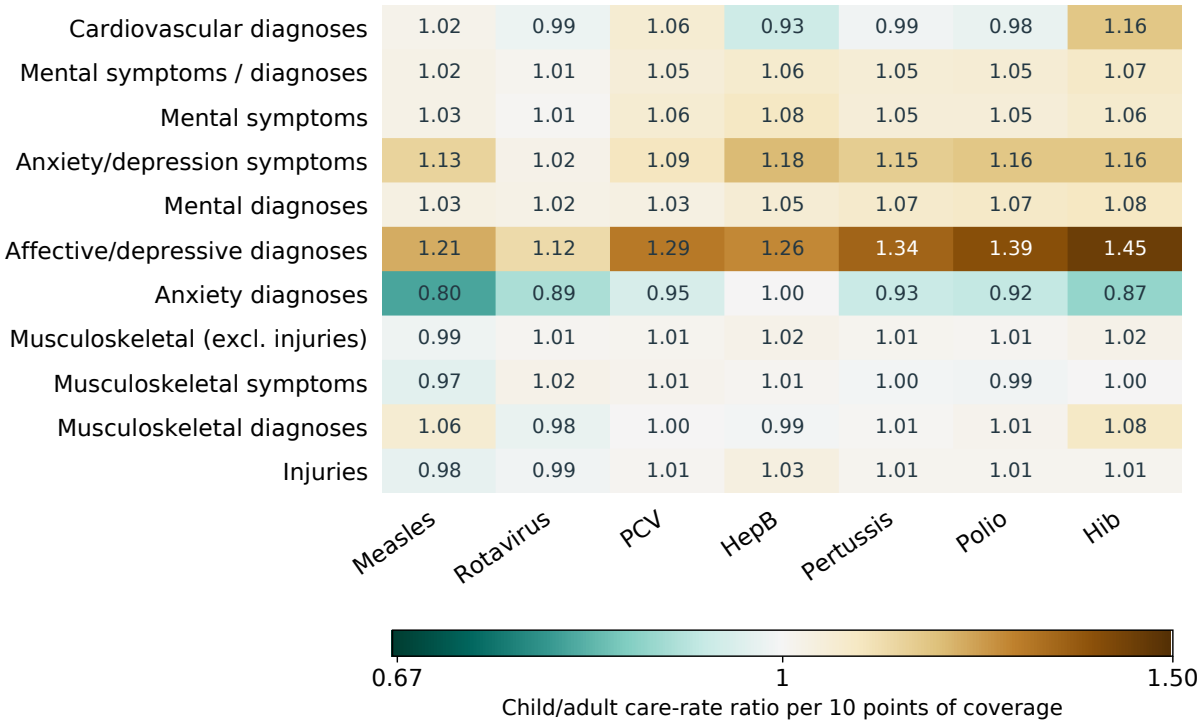


Figure S63. All 77 primary Norwegian age-gradient point estimates. Each number is a ratio of childhood/adult primary-care rate ratios per ten percentage points of age-two area coverage, with ages 45-64 as the comparison. Colors show direction and magnitude, not significance. None survives the stated 77-test adjustment. County-clustered intervals and power calculations are in the companion primary-results table. Outcomes and vaccine proxies overlap; these are not individual causal effects or autism-specific incidence estimates.

Conditional fixed-standard-error, noncentral-t calculations give at least 80% nominal power for a 10% change in the age-gradient ratio in 30 of 77 primary tests, and for a 25% change in 56 of 77. These refer to a ten-point area-coverage gradient, not individual vaccine harms. They use an unadjusted two-sided 5% test and do not incorporate multiplicity, confounding or failure of the comparison-age counterfactual. A ten-point coefficient unit is also not a feasible upward intervention at every observed coverage level. Every interval, p-value, conditional power and minimum-detectable-effect calculation is in the companion primary-results table.

The complete held-input rerun reproduces all 4,851 settings, comparing 103,994 numerical values with zero observed difference. Noncentral-t power quadrature is checked against the distribution calculation and a twice-dense quadrature. The central results register includes the complete Norwegian grid and unestimable statuses, with one shared outcome-system overlap label. These models are not added to the autism hypermean: they address primary-care age gradients under different observation and exposure assumptions. The source queries, metadata, frozen inputs, replay command and all estimates are available in `norway_primary_care`.

S59. Global integration, checked designs and shared-mean synthesis

S59.1 Source coverage

Every one of the 66 international crosswalk files has a disposition in `global_integration/dispositions.csv`. Two files receive two dispositions: eleven records are implemented autism sources, thirteen other-outcome sources, six replacements, sixteen overlapping populations, eighteen without usable timing variation and four blocked by exposure, outcome or input limitations. A disposition is not a new dataset or proof of worldwide acquisition completeness. A blocked or unestimable file is not a null result.

The adapters represent 101 blocks: 36 autism and 65 other-outcome blocks. Seven international autism blocks enter the current hierarchy as replacements or additions; 29 autism blocks remain reported without pooling. Other conditions have separate models. Product, exposure and kernel variants frequently reuse the same cells.

S59.2 Model and estimand

The restricted comparison model in this section includes fifteen pooled vaccine-location contexts in fourteen original likelihood blocks, comprising 5,133 outcome cells. U.S. PCV and RotaShield share a joint likelihood with 1,469 cells. U.S. rotavirus 2006 remains estimated in that likelihood but is untied from the shared mean and modern-rotavirus product. Japan contributes nine MMR cohorts. Other contexts cover Sweden, Germany, Hong Kong, England, Scotland, Northern Ireland, Canada excluding Quebec, Quebec, Czechia, Brazil, Colombia and Tanzania.

The model retains product and context heterogeneity, reconstructed exposure, trend and serial-dependence kernels, and context-bias distributions. Replacements substitute for existing likelihoods; they do not multiply information from the same outcomes. Quebec is represented once. Tanzania's PCV13-plus-rotavirus introduction remains a bundled contrast, not separate evidence for both products.

The exponentiated shared mean log effect is 0.957 (95 percent credible interval 0.599-1.517). It is a conditional mean over the included products and settings, not a population schedule-risk ratio or a directly observed vaccinated-versus-unvaccinated contrast. Context posteriors are likelihood-scale mixture parameters combining unseparated context bias and heterogeneity, not debiased individual causal risk ratios. The context, product and shared-mean marginals reuse the same observations.

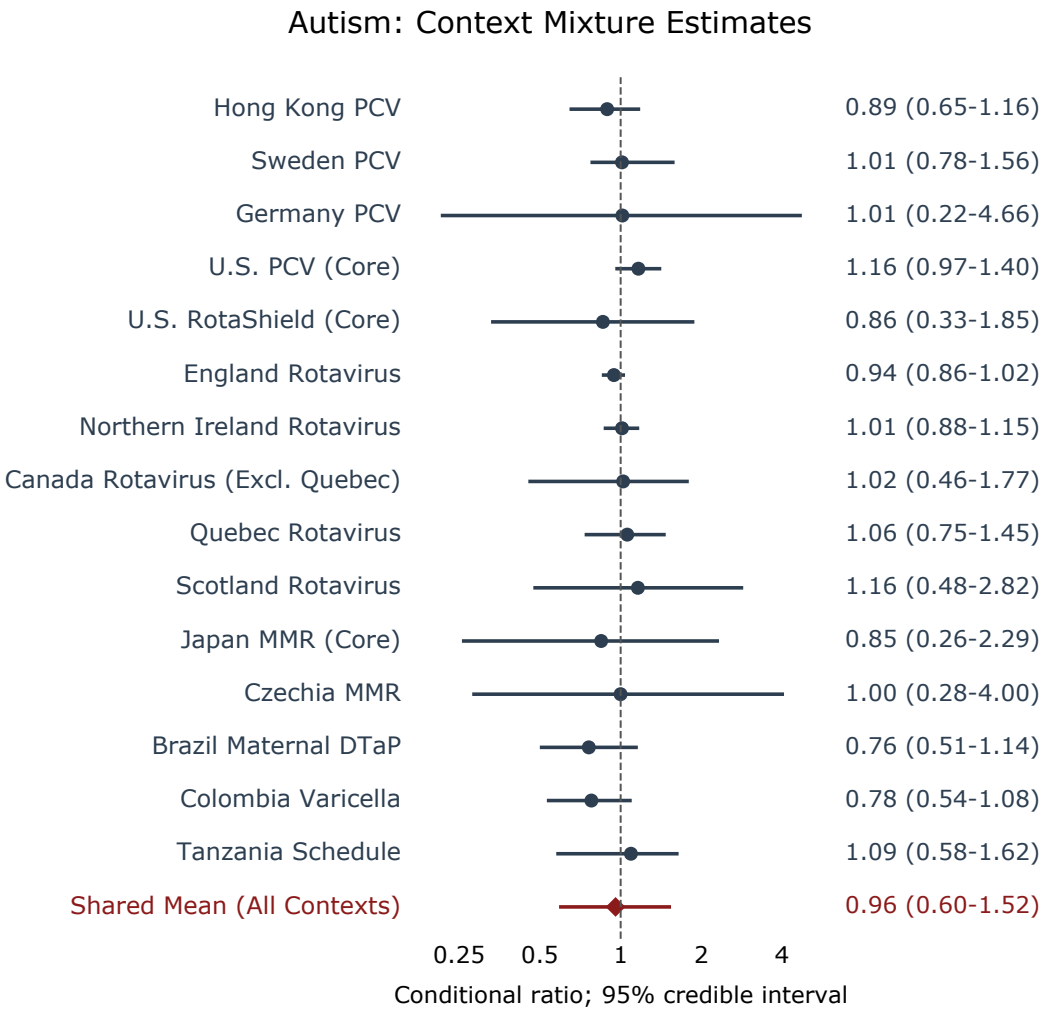


Figure S64. Context-mixture estimates and conditional shared mean. Context parameters include unseparated context bias and are not debiased individual causal risk ratios. Fifteen pooled contexts use fourteen original likelihood blocks; the U.S. PCV and RotaShield contexts share observations. U.S. rotavirus 2006 remains in their likelihood but is untied from the shared mean. Context marginals and the shared mean are different parameters, not independent studies. Bars are 95 percent credible intervals.

S59.3 Pooling gates

The D4 rule requires an autism-specific outcome, no duplicated observations, a future-exposure placebo including one under every specified kernel, no failed pretrend check, and acceptable synthetic-injection behavior. The future placebo shifts exposure three cohorts later on identical cells. An untestable pretrend is not a passed pretrend. Synthetic outcomes at specified risk multipliers assess the estimator's response, not whether the actual counterfactual is valid.

Rules were specified during the project rather than externally preregistered. Intervals condition on the selected blocks and omit uncertainty from gate selection. A low-powered placebo can pass; correlated future exposure is a timing diagnostic, not randomized assignment. The plasmode gate also depends on the observed outcome level, so an alternative including its sole failures is retained.

U.S. rotavirus 2006 is not pool-ready: its future placebo passes only after a post-hoc reporting-break screen, the unscreened version fails, and a required checker verdict is absent. Keeping the effect untied adjusts the other U.S. effects without treating it as an additional pooled null. Screened-and-pooled alternatives remain sensitivities.

Four pooled contexts fail the stricter two-sided placebo family: U.S. RotaShield, Northern Ireland rotavirus, Canada excluding Quebec and Quebec. Excluding them gives 0.954 (0.570-1.596). The Bonferroni version excludes none and duplicates the primary.

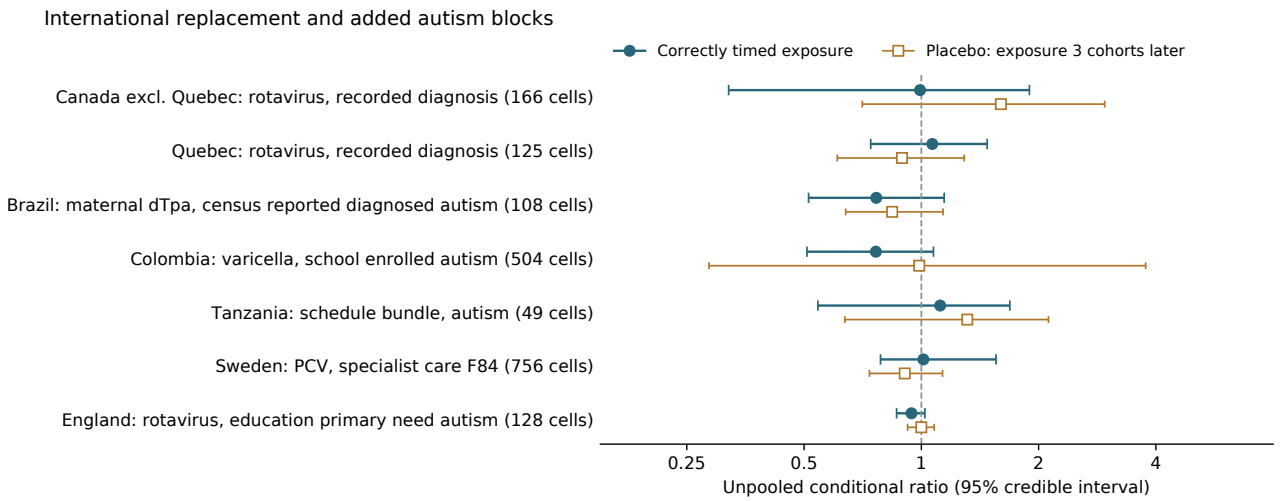


Figure S65. International autism blocks entering the global hierarchy. Filled circles are correctly timed unpooled conditional ratios with 95 percent credible intervals; open squares are the paired placebo with exposure shifted three cohorts later on identical cells. Replacement blocks substitute for, rather than duplicate, existing likelihoods. These are recorded-outcome ratios before synthesis bias uncertainty, not individual causal effects.

S59.4 Probabilities, sensitivity and numerical precision

The posterior probabilities that the shared mean exceeds 1, 1.10 and 1.25 are 0.425, 0.275 and 0.126. They do not describe whether any vaccine causes harm or every effect is zero. The continuous hierarchy assigns no point mass to a universal null. Both decreases and moderate increases remain compatible with this model.

Halving and doubling the shared-mean prior scale gives 0.965 (0.675-1.373) and 0.954 (0.478-1.893). Doubling context-bias standard deviations gives 0.974 (0.510-1.848). Including plasmode-only gate failures gives 0.896 (0.561-1.417). Keeping the reporting-break series with U.S. rotavirus untied gives 0.957 (0.599-1.517). These results remain assumption-dependent.

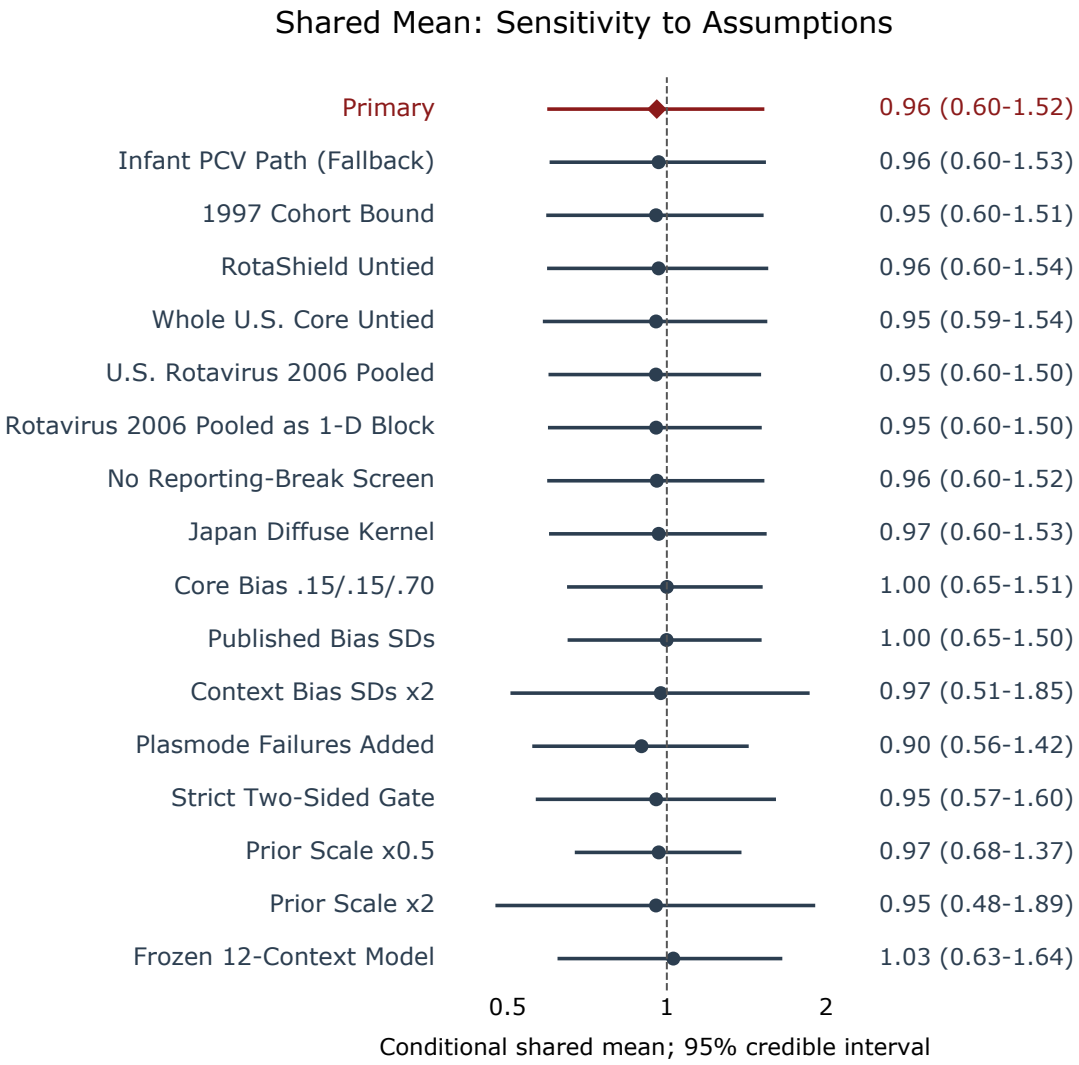


Figure S66. Shared-mean sensitivity to exposure, pooling and bias assumptions. Each row reuses the original observations. Bars are 95 percent credible intervals, conditional on the indicated model. The infant-path fallback is reported because its alternative joint fit retains low effective sample size. The full numerical flags and all remaining sensitivities are retained in the source report.

The primary uses power-18 Sobol integration, independent scrambles 23091 and 23092, and a 241-point shared-mean grid. “Power 18” denotes numerical draw count, not statistical power against harm. Its power-16 companion is 0.959 (0.601-1.520). The combined primary minimum relevant core effective sample size is 106. These are numerical checks, not identification tests.

Scenarios below the effective-sample-size floor of 50 are recomputed at power 20. Some retain low local effective sample sizes and remain flagged. The joint infant-PCV path has a minimum of 22.7 even at power 20; its specified fallback gives 0.965 (0.605-1.527), with a minimum of 2,318. The low-ESS alternative remains available. Removing an influential draw is a diagnostic, not the reported estimator.

S59.5 Nonpooled and other outcomes

The 29 nonpooled autism blocks include overlapping products, source variants and timing, trend, overlap or injection failures. Adverse- and protective-signed results remain visible; exclusion is not evidence of no effect. The 65 other-outcome blocks remain separate. Broader disability, care use and composition outcomes cannot be pooled as autism risk.

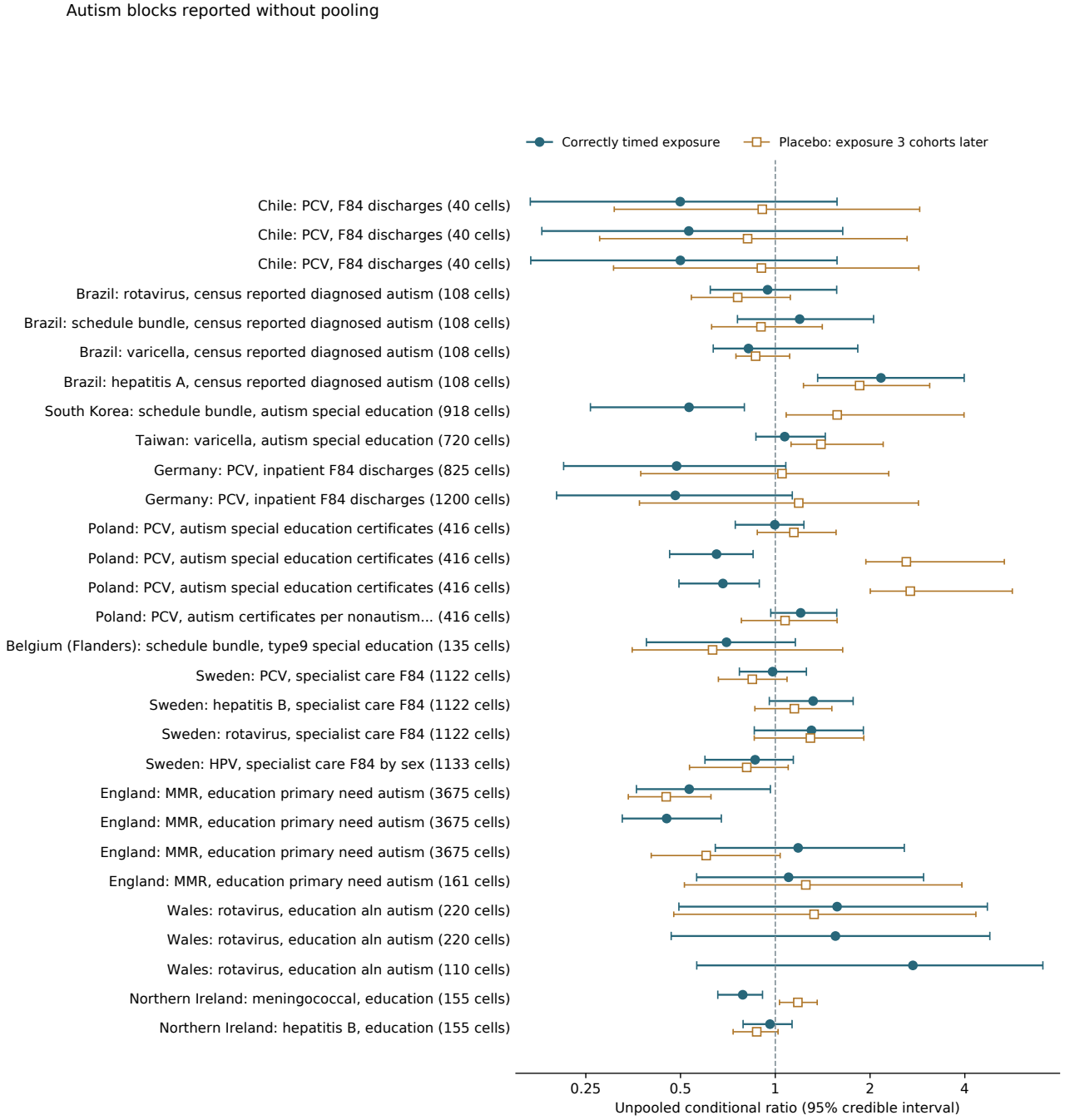


Figure S67. Autism blocks reported but not pooled. Rows are non-primary product variants on shared cells, sensitivities, and blocks failing a placebo, pretrend or overlap gate. A placebo that reproduces an association in either direction shows that the association is not timing-specific. Rows sharing cells are not independent.

Swedish rotavirus gastroenteritis is a positive-control check, not a result for every disability. Failed or imprecise positive controls limit validation. Another disability is not automatically a negative control: both the substantive claim and a credible absence of an exposure pathway must be specified. Outcome intervals are marginal and do not correct for the entire condition search.

Table S59-A. All 29 nonpooled autism blocks and the 12 of 65 other-outcome blocks whose 95 percent interval excludes one.

Block (class: place, product; outcome; identifier, cells)	Estimate (95% CrI)	Future +3 placebo (95% CrI)	Placebo and pretrend gates	D4 reason for not pooling
Autism: Chile, PCV; F84 discharges (CHL_PCV_F84, 40 cells)	0.50 (0.17-1.57)	0.91 (0.31-2.88)	Placebo passes (all 4 kernels); pretrend passes	Plasmode; reported as a hierarchy sensitivity
Autism: Chile, PCV; F84 discharges (CHL_PCV_F84_discharge_share, 40 cells)	0.53 (0.18-1.64)	0.82 (0.28-2.62)	Placebo passes (all 4 kernels); pretrend passes	Not a primary or replacement block
Autism: Chile, PCV; F84 discharges (CHL_PCV_F84_pop_base2024, 40 cells)	0.50 (0.17-1.57)	0.90 (0.31-2.86)	Placebo passes (all 4 kernels); pretrend passes	Not a primary or replacement block
Autism: Brazil, rotavirus; census reported diagnosed autism (BRA_RV_census_autism, 108 cells)	0.95 (0.62-1.57)	0.76 (0.54-1.12)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: Brazil, schedule bundle; census reported diagnosed autism (BRA_SCHED2010_census_autism, 108 cells)	1.19 (0.76-2.05)	0.90 (0.63-1.41)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: Brazil, varicella; census reported diagnosed autism (BRA_MMRV_census_autism, 108 cells)	0.82 (0.64-1.83)	0.87 (0.75-1.11)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: Brazil, hepatitis A; census reported diagnosed autism (BRA_HEPA_census_autism, 108 cells)	2.17 (1.36-3.98)	1.85 (1.23-3.09)	Placebo fails (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: South Korea, schedule bundle; autism special education (KOR_SCHEDULE_autism, 918 cells)	0.53 (0.26-0.80)	1.57 (1.08-3.98)	Placebo fails (all 4 kernels); pretrend fails	Placebo, pretrend, harm injection, plasmode
Autism: Taiwan, varicella; autism special education (TWN_VARICELLA_autism, 720 cells)	1.07 (0.87-1.44)	1.40 (1.12-2.20)	Placebo fails (all 4 kernels); pretrend passes	Placebo
Autism: Germany, PCV; inpatient F84 discharges (DEU_PCV_rebuilt, 825 cells)	0.49 (0.21-1.08)	1.05 (0.37-2.29)	Placebo passes (all 4 kernels); pretrend passes	Plasmode; reported as a hierarchy sensitivity
Autism: Germany, PCV; inpatient F84 discharges (DEU_PCV_rebuilt_16L_cc, 1200 cells)	0.48 (0.20-1.13)	1.19 (0.37-2.85)	Placebo passes (all 4 kernels); pretrend passes	One likelihood per series, plasmode
Autism: Poland, PCV; autism special education certificates (POL_PCV_SIO_autism_enrol, 416 cells)	1.00 (0.75-1.23)	1.15 (0.88-1.56)	Placebo passes (all 4 kernels); pretrend fails	Pretrend
Autism: Poland, PCV; autism special education certificates (POL_PCV_SIO_autism_gpop, 416 cells)	0.65 (0.46-0.85)	2.61 (1.94-5.34)	Placebo fails (all 4 kernels); pretrend fails	Not a primary or replacement block
Autism: Poland, PCV; autism special education certificates (POL_PCV_SIO_autism, 416 cells)	0.68 (0.49-0.89)	2.68 (2.00-5.66)	Placebo fails (all 4 kernels); pretrend fails	Not a primary or replacement block
Autism: Poland, PCV; autism certificates per nonautism certificate (POL_PCV_SIO_autism_shareoffset, 416 cells)	1.20 (0.97-1.57)	1.07 (0.78-1.57)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: Belgium (Flanders), schedule bundle; type9 special education (BEL_FL_SCHEDULE_type9, 135 cells)	0.70 (0.39-1.16)	0.63 (0.35-1.64)	Placebo passes (all 4 kernels); pretrend not tested	Not a primary or replacement block
Autism: Sweden, PCV; specialist care F84 (SWE_F84_allages_corrected, 1122 cells)	0.98 (0.77-1.26)	0.85 (0.66-1.09)	Placebo passes (all 4 kernels); pretrend not tested	One likelihood per series
Autism: Sweden, hepatitis B; specialist care F84 (SWE_HEPB_F84_allages, 1122 cells)	1.32 (0.96-1.77)	1.15 (0.86-1.51)	Placebo passes (all 4 kernels); pretrend not tested	One likelihood per series
Autism: Sweden, rotavirus; specialist care F84 (SWE_RV_F84_allages, 1122 cells)	1.30 (0.86-1.90)	1.29 (0.86-1.91)	Placebo passes (all 4 kernels); pretrend not tested	One likelihood per series
Autism: Sweden, HPV; specialist care F84 by sex (SWE_HPV_F84_sex, 1133 cells)	0.86 (0.60-1.14)	0.81 (0.53-1.10)	Placebo passes (all 4 kernels); pretrend not tested	One likelihood per series
Autism: England, MMR; education primary need autism (ENG_MMR_UTLA, 3675 cells)	0.53 (0.36-0.96)	0.45 (0.34-0.62)	Placebo fails (all 4 kernels); pretrend not tested	Placebo, plasmode
Autism: England, MMR; education primary need autism (ENG_MMR_UTLA_mmr_only, 3675 cells)	0.45 (0.33-0.67)	n/a	No placebo twin; pretrend not run	Not a primary or replacement block

Block (class: place, product; outcome; identifier, cells)	Estimate (95% CrI)	Future +3 placebo (95% CrI)	Placebo and pretrend gates	D4 reason for not pooling
Autism: England, MMR; education primary need autism (ENG_MMR_UTLA_utla_trends, 3675 cells)	1.18 (0.64-2.57)	0.60 (0.40-1.04)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: England, MMR; education primary need autism (ENG_MMR_NAT, 161 cells)	1.10 (0.56-2.96)	1.25 (0.51-3.91)	Placebo passes (all 4 kernels); pretrend not run	Not a primary or replacement block
Autism: Wales, rotavirus; education aln autism (WAL_RV, 220 cells)	1.57 (0.49-4.72)	1.33 (0.48-4.34)	Placebo passes (all 4 kernels); pretrend fails	Pretrend
Autism: Wales, rotavirus; education aln autism (WAL_RV_aln_uptake_break, 220 cells)	1.55 (0.47-4.80)	n/a	No placebo twin; pretrend not run	Not a primary or replacement block
Autism: Wales, rotavirus; education aln autism (WAL_RV_pre_aln, 110 cells)	2.74 (0.56-7.08)	n/a	No placebo twin; pretrend not run	Not a primary or replacement block
Autism: Northern Ireland, meningococcal; education (NIR_MENB, 155 cells)	0.79 (0.66-0.91)	1.18 (1.03-1.36)	Placebo fails (reference, spline-only, rho 0.85); pretrend not run	Not a primary or replacement block
Autism: Northern Ireland, hepatitis B; education (NIR_HEPB, 155 cells)	0.96 (0.79-1.13)	0.87 (0.73-1.02)	Placebo fails (linear-only); pretrend not run	Not a primary or replacement block
Other outcome: Poland, PCV; certificates intellectual mild (POL_PCV_SIO_INTELLECTUAL_MILD, 416 cells)	0.33 (0.17-0.57)	2.03 (1.39-3.71)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates intellectual mild (POL_PCV_SIO_INTELLECTUAL_MILD_enrol, 416 cells)	0.58 (0.36-0.77)	0.70 (0.45-1.06)	Placebo passes (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates intellectual moderate severe (POL_PCV_SIO_INTELLECTUAL_MODERATE_SEVERE, 416 cells)	0.48 (0.30-0.71)	2.03 (1.40-3.70)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates intellectual moderate severe (POL_PCV_SIO_INTELLECTUAL_MODERATE_SEVERE_enrol, 416 cells)	0.71 (0.47-0.94)	0.69 (0.48-0.99)	Placebo fails (reference, linear-only, rho 0.85); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates multiple disabilities (POL_PCV_SIO_MULTIPLE_DISABILITIES, 416 cells)	0.47 (0.31-0.64)	2.70 (1.99-5.78)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates multiple disabilities (POL_PCV_SIO_MULTIPLE_DISABILITIES_enrol, 416 cells)	0.69 (0.51-0.83)	1.03 (0.81-1.34)	Placebo passes (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates hearing (POL_PCV_SIO_HEARING, 416 cells)	0.54 (0.35-0.80)	2.38 (1.67-4.62)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Poland, PCV; certificates vision (POL_PCV_SIO_VISION, 416 cells)	0.39 (0.16-0.75)	1.66 (1.08-2.92)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: France, hepatitis B; cnam childhood onset psych (FRA_HEPB_PSY_ENF, 2484 cells)	0.55 (0.44-0.67)	0.51 (0.41-0.64)	Placebo fails (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: France, hepatitis B; cnam epilepsy (FRA_HEPB_EPI, 2538 cells)	0.86 (0.76-0.98)	0.94 (0.80-1.09)	Placebo passes (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)
Other outcome: France, hepatitis B; cnam other neurologic (FRA_HEPB_NEU_AUT, 1530 cells)	1.46 (1.17-1.81)	1.37 (1.04-1.76)	Placebo fails (reference, linear-only, spline-only); pretrend not run	Outcome (class other; separate posterior)
Other outcome: Sweden, rotavirus; specialist care A08 (SWE_RV_A08, 288 cells)	0.46 (0.19-0.95)	0.43 (0.16-1.09)	Placebo passes (all 4 kernels); pretrend not run	Outcome (class other; separate posterior)

Notes. Estimates are unpooled conditional ratios under the reference kernel and prior, with 95 percent credible intervals; the placebo shifts exposure three cohorts later on identical cells. Where a placebo fails, the kernels (reference, linear-only, spline-only, rho 0.85) whose placebo interval excludes one are listed. Not run marks non-candidate blocks whose gate evaluation stopped before that component; not tested marks blocks without a separate pretrend test, which is

not a passed pretrend. n/a: no future +3 placebo twin. The Welsh rotavirus pretrend failure is a single pre-period lead (p = .010), and the window spans the 2021 change in the additional-learning-needs system. Polish rows differ in the offset: live births (no suffix), population (gpop), enrolment (enrol, the primary design) or non-autism certificates (shareoffset). The Swedish hepatitis B, rotavirus and HPV rows are alternative likelihoods on the Swedish PCV F84 series (the same patients); they are product-attribution sensitivities, not additional evidence. Other-outcome blocks have separate posteriors and never enter the autism model. Intervals are marginal and unadjusted for the number of blocks examined. Source: `global_integration/blocks.csv` and `global_integration/gate_decisions.json`.

S59.6 Checked designs and source units

The master inventory retains 206 rows: 57 jump rows, 110 interaction rows and 39 synthetic-control rows. The jumps comprise 56 inventory events and a separate NHIS series; 37 have a checked estimate and 31 a calibrated interval. Interactions comprise 105 pairs and five single margins. Synthetic controls comprise 37 designs and two comparison/control rows, principally diagnostics rather than independent likelihoods.

Original source rows, first stages, interval construction and coverage, harm-direction flags, checker findings and overlap groups remain attached. Some synthetic-control intervals have coverage other than 95 percent; the register preserves it. Missing conversions and unestimated effects stay missing. The cumulative schedule-load exercise lacks an identified estimand; that is not evidence of a zero load effect. All 206 master rows have a separate register collection. Numerical reruns and exact aliases of hierarchy scenarios are not counted again.

ADDM's six recurring sites have an early-PCV uptake standard deviation of 2.35 percentage points, versus 8.91 points across IDEA states. ADDM's 1.053 (0.971-1.143) per six-site SD becomes 1.247 (0.881-1.766) per ten points under the log-linear transformation. The ADDM-to-IDEA comparison on the common ten-point scale is 1.230 (0.941-1.606), p=0.130. Nominal ADDM power for a ratio of 1.10 per ten points is about 8.9 percent under the stated calculation, not its larger power per native SD. The observed site range is only about 7.2 points, so ten points also requires extrapolation. None of these native-scale calculations validates the withdrawn individual PCV mapping.

S59.7 Source files, numerical sensitivity and reproducibility

The authoritative source is `global_integration/hierarchy/hierarchy_final_p18.json`, with its power-16 companion and source-hashed effective-sample-size diagnostics. `3320_final_synthesis.csv` and its report list every scenario and marginal target. `3309_jump_master_table.csv` and the harm-direction register retain the checked designs. The source-unit repair is documented in `3240_addm_unit_and_label_repairs.py` and its audit.

The global validator checks dispositions, block curves, gates, overlap, numerical comparisons and current report hashes. Register checks cover exact source endpoints, specified interval coverage, every master row, every specified posterior summary and workbook completeness. These are computational checks, not a causal-identification certificate.

The November-2005 onset sensitivity replaces only the 2005 U.S. rotavirus exposure point, reducing the state mean from 3.58 to 0.54 percent and preserving every other cohort. Existing bootstrap draws retain their draw-to-point ratios; this is a hybrid timing sensitivity, not a full reconstruction of checker-bootstrap uncertainty. The U.S.-alone rotavirus ratio changes from 0.952 (0.757-1.192) to 0.928 (0.748-1.155). Because this context remains untied, the shared mean changes only from 0.9572 (0.5988-1.5170) to 0.9576 (0.5993-1.5173); its minimum relevant effective size is 148. This does not clear the context's pooling gates.

The infant-PCV and Japan-only-bias sensitivities retain minimum effective sizes of 22.7 and 9.3 after power-20 runs. Influential draws are retained; close seed-specific medians do not validate tail probabilities. The specified infant fallback is interpretive, while the Japan-only-bias setting remains numerically unsettled. A different, independently validated proposal or quadrature method is required before these settings can support precise bounds. Four source dispositions remain unsuitable or blocked rather than contributing null observations.

Fresh package extraction verified 2,627 manifest entries in the tested snapshot, reproduced 936 international numerical values exactly, and rebuilt the two power-18 U.S./Japan likelihoods and current reference shared-mean result with exact agreement in seven checked fields. The final package manifest identifies its current membership. This is held-input current-model reproduction, not a claim that every historical analysis has been rebuilt from newly downloaded raw files. The isolated execution receipts and commands accompany the package; limitations in acquisition, identification and numerical sensitivity remain explicit.

S60. The jump program

S60.1 Inventory and feasibility

A jump is a sharp, dated change in the share of a birth cohort vaccinated: an introduction, withdrawal, deferral or eligibility change. The inventory lists 56 candidate uptake jumps: 27 in the United States, 10 in the United Kingdom, 6 in Japan, 2 in Sweden, one each in New Zealand, Chile, Latvia, Canada, Spain, Germany, Denmark, France, Italy and Taiwan, and one multi-country entry for the remaining screened events. Five were rated good, 23 marginal and 28 infeasible for a coverage-instrumented fixed-age design. Ten jumps went to a first estimation round and 40 to an expansion round in eight batches; six were never sent to either round, and one of those, the Hib booster shortage, was estimated separately (Section S60.4). An adversarial checker reviewed all 50 jumps that were sent, confirming, correcting or rejecting each estimator, and judged 15 of them not estimable from public data.

Thirty-nine of the 56 first-stage sizes are rebuilt from project data, 11 are unmeasured, one is not an uptake jump (thimerosal reduction changes dose, not uptake) and five rest on assumed or proxy values; 281 checked numbers reproduce. The assumed values include the 89.5 percentage points the inventory lists for England's 2006 PCV introduction, which is in fact the England rotavirus step; the checker replaced it with a measured HPA COVER coverage path, whose step from the 2004 to the 2007 cohort is +89.4 points, and Table S60-A carries that measured value, which the inventory's reproduction (file 3300) did not recompute. They also include the roughly 90-point values for Northern Ireland, which are schedule-eligibility shares rather than measured uptake. U.S. inventory first stages use a survey-year proxy cohort map that misplaces PCV7 exposure by cohort; the corrected any-dose and infant paths are in the PCV7 cohort-exposure file (3200).

S60.2 Conversion, calibration and checking

Each jump is fitted as $E[y_c] = N_c \exp(s(c)) [1 + p_c (r - 1)]$, where N_c is the cohort denominator, $s(c)$ the specified counterfactual path, p_c the measured share of cohort c exposed and r the rate ratio of a vaccinated relative to an unvaccinated child. Intervals are calibrated against placebo breakpoints: the coverage path is shifted three or more cohorts earlier within pre-exposure data, and the calibrated standard error is the largest of the model standard error and the standard deviation and root mean square of the placebo log r. A few families use variants (drop-one-cohort jackknife standard errors in Northern Ireland, negative-control standard errors in Sweden, site-cluster standard errors for ADDM, and placebo shifts of two to four cohorts for the Hib shortage, whose national design has only three placebos). Where the checker corrected an estimator, the corrected value is the headline. An interval is uncalibrated where fewer than three admissible placebos exist or the family's rule could not calibrate it. Several calibrated intervals are fallbacks for designs in which no specification passed its placebo check, so "calibrated" does not always mean "placebo-passing".

The checked master table holds 206 rows: 57 jump rows, 110 interaction rows (105 pairs and five single-vaccine margins) and 39 synthetic-control rows (37 designs and two negative controls). The 57 jump rows are the 56 inventoried jumps with the NHIS part of the U.S. survey jump carried separately. Thirty-seven have a checked estimate and 31 a calibrated interval. A jump's own row never enters the shared mean. Jumps whose outcome series is already a hierarchy context (3 core, 6 pooled and 6 report-only contexts) are validation only; five enter a likelihood only as nuisance columns (three in the core U.S. likelihood and two in another block), two are same-cell proxies for the untied U.S. rotavirus 2006 context, one is retired, and 34 are excluded, most because they fail the reconciliation's uniform gate. Overlap groups partition blocks and contexts so that taking one likelihood per group never reuses cells.

S60.3 Every jump on the individual scale

None of the 37 checked estimates has a headline interval that excludes one in either direction. Among the 31 calibrated estimates, 16 lie above one and 15 below. The most precise come from large national introductions: England PCV 2006, 0.941 (0.862-1.027); England rotavirus 2013, 0.951 (0.825-1.097); and U.S. PCV7 in the IDEA state panel, 1.042 (0.883-1.230). The England PCV estimate carries three caveats. It measures the whole September 2006 package rather than PCV alone, because a Hib/MenC booster began at the same time and the two cohort paths are correlated 1.0000. It is outside the pool because it uses the same national DfE cells as the pooled England rotavirus block. And the same design applied to pupils with speech, language and communication needs gives r of 0.899-0.969 across its ten specifications, with nine of the ten profile intervals below one, so the dip is not autism-specific. For England rotavirus no specification passes its placebo check (six of seven placebo intervals exclude one), so the widest calibrated interval is reported. Four of the jump estimates behind pooled contexts carry no information, for the reasons the checker recorded: the RotaShield jump is small (8.7 points) and no specification passes calibration; the Scottish estimate rests on vacuous placebos and duplicates the pooled Scottish block, and Wales is not estimable; no Swedish PCV specification passes placebo calibration; and German PCV cannot be separated from MenC (correlation -0.95). Twenty rows have no estimate: the checker confirmed fifteen as not estimable, among them the Japanese special-class jumps and the Canadian, Chilean, Latvian, New Zealand and Spanish rollouts, and five were never sent for estimation. The 37 estimates include the California DDS design, which the checker judged invalid, and the Yokohama fit, which carries no finite calibrated information.

Table S60-A. All 57 jump rows of the checked master table.

Jump and outcome series	First stage (pp)	r (95% interval)	Calibrated	Status
U.S. PCV7 2000, IDEA state panel	87.9	1.042 (0.883-1.230)	yes	core context
U.S. PCV7 2000, California CASEMIS	91.6	1.027 (0.569-1.853)	yes	excluded
U.S. PCV7 2000, Minnesota districts	87.5	1.228 (0.233-6.484)	yes	excluded
U.S. PCV7 2000, New Jersey districts	92.0	0.542 (0.051-5.815)	no	excluded
U.S. PCV7 2000, New Hampshire	92.4	1.279 (0.484-3.382)	yes	excluded
U.S. PCV7 2000, ADDM sites	87.9	0.900 (0.604-1.341)	no	excluded
U.S. PCV7 2000, IDEA national	87.9	0.873 (0.475-1.605)	yes	excluded
U.S. PCV7 2000, NSCH	87.9	1.926 (0.432-8.580)	yes	excluded
U.S. PCV7 2000, NHIS	87.9	0.881 (0.280-2.769)	yes	excluded
U.S. RotaShield 1998-1999, IDEA age 8	8.7	0.843 (0.278-2.554)	yes	core context
U.S. RotaShield 1998-1999, ADDM sites	8.7	1.127 (0.043-29.409)	yes	excluded
U.S. rotavirus 2006, IDEA ages 3-5	56.1	1.313 (0.739-2.335)	yes	same-cell proxy (validation)
U.S. rotavirus 2006, California CASEMIS	58.0	1.216 (0.581-2.547)	yes	same-cell proxy (validation)
U.S. rotavirus 2006, Minnesota districts	56.8	0.939 (0.664-1.328)	yes	retired
U.S. rotavirus 2006, New Hampshire	61.6	0.944 (0.328-2.717)	yes	excluded
U.S. rotavirus 2006, ADDM sites	56.1	not estimable	-	excluded
U.S. universal hepatitis A 2006	46.6	1.055 (0.470-2.367)	yes	excluded
U.S. targeted hepatitis A 1999	35.0	1.119 (0.690-1.815)	yes	core nuisance column
U.S. varicella 1995-1996	45.3	1.092 (0.793-1.502)	yes	excluded
U.S. influenza, 6-23 months, 2004	50.5	1.131 (0.856-1.495)	yes	core nuisance column
U.S. infant hepatitis B 1991	-	0.898 (0.697-1.156)	yes	excluded
U.S. thimerosal reduction 1999-2001	-	not estimated	-	excluded
U.S. second MMR dose 1989-1990	-	0.823 (0.665-1.012)	no	excluded
U.S. Hib booster shortage 2007-2009	-7.2	1.055 (0.717-1.553)	yes	core nuisance column
U.S. state rotavirus mandates 2011-2021	-	not estimable	-	excluded
Japan MMR withdrawal, Yokohama Kohoku	-69.8	2.075 (0.000-4.3 million)	no	core context
Japan Hib and PCV 2008-2013, MEXT classes	-	not estimable	-	excluded
Japan routine varicella 2014, MEXT classes	-	not estimable	-	excluded

Jump and outcome series	First stage (pp)	r (95% interval)	Calibrated	Status
Japan routine hepatitis B 2016, MEXT classes	-	not estimable	-	excluded
Japan encephalitis suspension 2005 and reinstatement 2010	72.6	not estimable	-	excluded
Japan rotavirus 2020 and HPV suspension 2013	-67.0	not estimable	-	excluded
England rotavirus 2013	88.3	0.951 (0.825-1.097)	yes	pooled context (validation)
England MenB 2015	92.5	0.858 (0.545-1.351)	yes	nuisance column
England hexavalent hepatitis B 2017	91.8	1.126 (0.819-1.549)	yes	nuisance column
England PCV 2006	89.4	0.941 (0.862-1.027)	yes	excluded
England MMR decline 1998-2003	-11.9	not estimable	-	report-only context
England local-authority MMR decline 2012-2020	-3.8	not estimated	-	report-only context
Northern Ireland rotavirus 2013	90.0	0.914 (0.729-1.146)	yes	pooled context (validation)
Northern Ireland MenB 2015	90.0	0.895 (0.665-1.205)	yes	report-only context
Northern Ireland hexavalent hepatitis B 2017	90.0	1.088 (0.646-1.833)	yes	report-only context
Scotland and Wales rotavirus 2013	91.0	0.824 (0.253-2.686)	yes	pooled context (validation)
Sweden PCV 2007-2009	98.0	1.101 (0.251-4.827)	yes	pooled context (validation)
Sweden rotavirus 2014-2019	83.0	0.909 (0.452-1.825)	yes	report-only context
New Zealand PCV, rotavirus and varicella	87.0	not estimable	-	excluded
Chile PCV10 2011 and other events	-	not estimable	-	report-only context
Latvia PCV 2010 and rotavirus 2015	76.0	not estimable	-	excluded
Canada provincial rotavirus 2010-2020	-	not estimable	-	pooled context (validation)
Spain varicella 2016	-	not estimable	-	excluded
Germany PCV and MenC 2006	68.4	0.655 (0.020-21.462)	yes	pooled context (validation)
Denmark PCV 2007	91.0	not estimated	-	excluded
France mandate extension 2018 (psychiatric outcome)	18.8	0.893 (0.699-1.140)	yes	excluded
Italy schedule 2017	16.3	not estimated	-	excluded
Taiwan PCV13 2015 and other events	-	not estimable	-	excluded
Other international events, blocked	-	not estimated	-	excluded
U.S. PCV7 and rotavirus, Massachusetts IEPs	87.1	not estimable	-	excluded
U.S. PCV7 2000, New York SEDCAR	91.9	0.835 (0.569-1.260)	no	excluded
U.S. PCV7 2000, California DDS	91.6	0.766 (0.414-1.408)	no	excluded

Notes. r is the individual-scale rate ratio of a vaccinated relative to an unvaccinated child, with its headline 95 percent interval (all 37 estimated rows have nominal 95 percent coverage). Calibrated: yes if the interval is placebo-calibrated under the family's rule. First stages are the inventory's values in percentage points, except England PCV 2006, which carries the checker's measured +89.4 points on the HPA COVER path in place of the inventory's assumed 89.5, the England rotavirus step (Section S60.1); negative values are withdrawals or deferrals, and a dash marks an unmeasured or non-uptake first stage. The Hib shortage row lists the inventory's three-dose change; its estimation uses the booster-by-19-months dip of 50.0 points (Section S60.4). The Northern Ireland first stages are assumed values (Section S60.1). England PCV 2006 shares the national DfE cells of the pooled England rotavirus block, its r is the whole September 2006 package (PCV and the Hib/MenC booster), and the same design returns a dip as large for speech, language and communication needs (Section S60.3). Status describes the row's role: a pooled or core context is validated, not counted again. Source: `3309_jump_master_table.csv` data rows 1-57, `3300_jump_inventory.md`, `3301_jump_first10.md`, `3302_jump_expansion_b1_b4.md` and `3303_jump_expansion_b5_b8.md`.

S60.4 Deferral: the Hib booster shortage

The 2007-2009 Hib shortage deferred the booster dose. It is a timing contrast, a booster by 19 months versus late or never, not Hib vaccination versus none; the primary series was unaffected. On-schedule booster receipt by 19 months fell to 27.9 percent in the 2007 cohort against a 2003-2006 mean of 77.9 percent, a dip of 50.0 percentage points (the checker's deconvolution gives 49.9), and by 14.5 points in the 2008 cohort. Other vaccines at the same visit barely moved (MMR1 -2.4, DTaP4 -1.6 and two or more infant Hib doses -1.5 points), and booster receipt by 30 months was 55.6 percent, so most deferred children were boosted late rather than never. Rotavirus uptake rose in the same cohort (19 percent in 2006, 69 percent in 2007), which is why the checker added joint rotavirus and hepatitis A terms.

The corrected national estimate is r = 1.055 (calibrated 0.717-1.553). The cross-state design relates each state's dip to its autism outcome: the coefficient is +0.00085 per ten-point shortfall (p = 0.86), and the individual-scale state model gives r = 0.992 (calibrated 0.811-1.212). Event-study leads for 2003-2006 do not reject (F = 1.09, p = 0.37), and the state dip is uncorrelated with state rotavirus (r = -0.020) and hepatitis A uptake (r = +0.049). Identification rests on one full-dip cohort, the outcome is autism classification at ages 12-17, and the national calibrated interval depends on three placebos and on the spline basis.

S60.5 The harm-direction register

The harm-direction register lists every stored estimate that points in the harmful direction under a specified rule, with its diagnostics. For jump-program estimates the rule is a stored interval entirely above one, or a native log gap above zero where no finite r exists. For hierarchy-context, U.S.-alone and unpooled block posteriors the rule is a posterior 95 percent lower bound above 0.95 or P(r > 1) of at least 0.9, so some of those rows have intervals that include one. The register holds 154 rows: 89 jump-program signals, one hierarchy context, seven U.S.-alone diagnostics, nine unpooled block posteriors, 39 project chronic and disability estimates and nine published chronic estimates. Scanning 1,787 estimate records gave the 89 jump-program signals in three tiers: 3 in a row's own headline interval (tier 1), 55 in another specification, variant or superseded run whose calibrated or otherwise final interval lies above one (tier 2), and 31 in an uncalibrated interval where no calibrated one exists, including in-space-only Canadian synthetic-control intervals (tier 3). A further 83 records have a model interval above one but a calibrated interval that includes one; under the calibration rule they are not signals.

Table S60-B. Jump-program harm-direction signals by family.

Family and diagnostics	Tier	Signal range or estimate	Headline or context
Influenza by Hib deferral interaction: BH q 0.354 across 38 primary rows; excess in a Hib main effect of 0.761 (0.628-0.923) among children without influenza vaccine; coincides with the 2009 H1N1 season and the Hib shortage; time-shift placebos do not reproduce it	1 (checked value); 2 (errors-in-variables value)	1.373 (1.085-1.737); errors-in-variables 1.501 (1.110-2.030)	pair in the 27-pair layer
Swedish municipal rotavirus synthetic control, Jönköping: outside the donor hull; gap present before exposure; outcome is support recipients, not autism	1	4.702 (1.128-9.029)	discarded
Swedish early-county PCV synthetic control: gap exceeds the ceiling attainable for any r and reappears under equal exposure	1	log gap 0.734 (0.590-0.870), no finite r	discarded
England hexavalent hepatitis B, 12 specifications: four of eight placebo intervals exclude one; timing falsification fails; injury deaths on the same national step 2.21 (1.17-3.94)	2	1.069-1.214	1.126 (0.819-1.549)
U.S. influenza, 6-23 months, 15 specifications: one of six placebo intervals excludes one; includes superseded first-round runs; nuisance column only	2	1.204-1.469	1.131 (0.856-1.495)
U.S. targeted hepatitis A 1999, 12 signals: one superseded calibrated value, 1.113 (1.029-1.203); the other 11 are co-vaccine nuisance coefficients (the PCV7 path inside the hepatitis A model, and the hepatitis A path inside a superseded influenza model), not checked estimates	2 and 3	1.110-1.509	1.119 (0.690-1.815)
Canadian provincial rotavirus synthetic controls, 21 variants, plus one uncalibrated province-by-year jump-design diagnostic: nominal coverage 0.50-0.89 or in-space-only intervals; small first stage (Section S61.3)	2 and 3	2.018-9.534; uncalibrated diagnostic 1.788 (1.312-2.367)	hierarchy 0.99 (0.32-1.89)
Swedish municipal synthetic-control variants, 7: donors with earlier hepatitis B or rotavirus programmes coded unexposed; national records contradict the hepatitis B coding	2	4.427-9.663	discarded or convergence-only

Family and diagnostics	Tier	Signal range or estimate	Headline or context
ADDM: no admissible PCV7 placebo; the interaction is theta-only and not identified; the rotavirus fit sits at its bound and was judged not estimable	2 and 3	nine uncalibrated PCV7 specifications 1.241-1.415; influenza by PCV13 4.443 (1.492-20.086); rotavirus 2006 3.000 (1.012-8.895)	kept out of the pool
MMR dose 2 by Hib deferral: sensitivity tier with imputed joint receipt	2	1.687 (1.048-2.717); errors-in-variables 2.089 (1.068-4.086)	1.120 (0.553-2.269)
New Hampshire PCV7: placebo breakpoints return log r of +0.6154 and +0.5981	2	1.631 (1.294-2.056)	1.279 (0.484-3.382)
France mandate extension 2018: psychiatric outcome, not autism-specific	2	1.174 (1.023-1.347)	0.893 (0.699-1.140)
U.S. universal hepatitis A 2006: uncalibrated	3	1.226 (1.042-1.443)	1.055 (0.470-2.367)
U.S. second MMR dose: uncalibrated; no placebo feasible	3	1.942 (1.624-2.310)	0.823 (0.665-1.012)

Notes. Tier 1 is a row's own headline interval; tier 2 another specification, variant or superseded run with a calibrated or otherwise final interval above one; tier 3 an uncalibrated interval. Family ranges combine the listed specifications. Source: `3309_harm_direction_register.md` and `.csv`, `3309_jump_master_table.md`.

The U.S. PCV context is the one pooled context that meets the posterior rule: 1.165 (0.970-1.395) as the pooled context posterior, which its 0.70 bias allowance leaves at the own-data value, with $P(r > 1) = 0.948$. Its U.S.-alone diagnostics run from 1.141 to 1.212, with the 1997-bound exposure at 1.212 (1.004-1.468); the infant-exposure sensitivity gives 1.456 (1.169-1.863). Its future-exposure placebo includes one under all four kernels (reference 0.99, 0.68-1.56), but its IDEA negative controls lean at least as far and the NSCH design shows a harm-direction bias signal in the same children (Section S63). The nine unpooled block posteriors are the Brazilian hepatitis A census variant, 2.166 (1.363-3.981), whose future+3 placebo is 1.85 (1.23-3.09); French hepatitis B coverage with other neurologic conditions, 1.457 (1.170-1.809), placebo 1.37 (1.04-1.76), and with chronic respiratory conditions, 1.147 (0.940-1.353), placebo 1.58 (1.28-1.95); and six posteriors whose intervals include one: a Polish PCV autism-share block, 1.204 (0.967-1.568); Swedish hepatitis B specialist care, 1.319 (0.957-1.767); two Swedish PCV non-autism outcomes, 1.078 (0.959-1.305) and 1.193 (0.995-1.411); Taiwanese hepatitis A developmental delay, 1.126 (0.956-1.437); and a Welsh rotavirus block, 2.736 (0.563-7.079). The 48 chronic and disability rows are discussed with their diagnostics in the main text's chronic section and in Sections S47-S51.

Every harm-direction row outside the chronic outcomes therefore belongs to one of five classes: a non-headline or uncalibrated specification of a jump whose checked interval includes one; a synthetic control discarded as invalid or kept for convergence only; a pair that fails multiplicity and coincides with a calendar confound; an estimate whose own placebo or negative controls return the same kind of "effect"; or an unpooled posterior whose interval includes one. The U.S. PCV lean belongs to the fourth class but remains unresolved, because only one of its IDEA controls is precise enough to count. Files: `3300_jump_inventory.md` and `.csv`; `3301_jump_first10.md`; `3302_jump_expansion_b1_b4.md`; `3303_jump_expansion_b5_b8.md`; `3304_hib_booster_deferral.md`; `3309_jump_master_table.md` and `.csv`; `3309_harm_direction_register.md` and `.csv`.

S61. Synthetic controls on the jump cells

S61.1 Role, counts and conversion

Synthetic controls and synthetic difference-in-differences re-estimate selected jumps on cells the hierarchy already uses. They are same-data convergence checks, not independent evidence blocks, and no synthetic-control row enters the shared mean. The checked master table holds 39 synthetic-control rows: 37 designs and two negative-control rows. Twenty are convergence-only, one is convergence-only with zero causal weight, one is report-only because its outcome is an autism composite, two are negative controls kept for convergence only, and fifteen are discarded (six from the U.S. and Canadian file and nine from the international file). A gap in the outcome is converted to the individual scale by the mixture relation through the treated units' first stage, so a small first stage magnifies a noisy gap. Intervals come from donor-placebo or test-inversion distributions whose nominal coverage depends on the number of placebo units; some are 90, 75 or 50 percent intervals, and the register preserves that coverage.

S61.2 United States

U.S. hepatitis A 1999 compares 15 targeted states with 25 donors; states with their own mandate or missing years are excluded. Treated any-dose coverage at 30-35 months was 45.1 percent in 2000 and 60.5 percent in 2003, against 1.0 percent in donors in 2000. The stacked synthetic difference-in-differences estimate for 2003-2011 is a log gap of +0.0437 (95 percent placebo interval -0.112 to +0.399), with a pre-period RMSPE of 0.0184 against a null median of 0.0286 and an absolute-effect placebo p of 0.625; with a first stage of 26.0 percentage points it converts to r = 1.175 (0.611-3.244). The dose-response synthetic difference-in-differences gives 1.23 (0.81-1.78), synthetic control 0.97 (0.47-1.51) and augmented synthetic control 1.18 (0.77-1.77). The checked jump on overlapping cells gives 1.12 (0.69-1.82). All lean above one with intervals that include one, and all synthetic controls are convergence-only. The outcome is autism's share of IDEA enrolment at ages 6-11, and autism was only 4.1-8.3 percent of that enrolment in 2005-2011, so a true rise is attenuated by roughly that share. The aggregate treated series lies inside the donor hull, so its synthetic-control r is not identified and is not reported.

State mandate stacks estimate log gaps without an identified r. Hepatitis A (14 states) gives -0.0161 (-0.072 to +0.044), PCV7 (28 states) -0.0386 (-0.084 to +0.009), PCV13 (9 states) +0.0248 (-0.039 to +0.089) and rotavirus (5 states) +0.0572 (-0.063 to +0.220). The Tdap negative control (11 states) gives +0.0316 (-0.054 to +0.265) but rejects on the stacked RMSPE-ratio test at p = 0.015, so that test is anti-conservative. Hepatitis B and varicella day-care mandates are not estimable under the strict support rules. Six stacked non-medical exemption repeals give +0.0257 (-0.045 to +0.079), whose maximum attainable gap of 0.0128 leaves r unbounded above; the eight exemption-event diagnostics have first stages from -1.263 to +4.0 percentage points and zero causal weight (Section S22).

S61.3 Canada

Canadian provincial rotavirus introductions use Chronic Disease Surveillance System autism counts at ages 1-19, with Prince Edward Island, Ontario, Quebec, British Columbia and Saskatchewan as treated provinces and Manitoba, Alberta and Nova Scotia as donors, in a clean window ending in fiscal year 2014. The headline demeaned synthetic control is a gap of +0.0746 on a treated exposure share of 0.0807, giving r = 1.960 (0.000-6.941) as a pooled-null 90 percent interval. Province and specification variants give 1.761 (0.000-6.424) for Ontario, 1.814 (0.000-7.051) for Quebec, 4.993 (0.000-12.199) for British Columbia, whose gap is already +0.25 in fiscal year 2011 when exposure is zero, 0.000 (0.000-1.250) for Saskatchewan, whose gap sits at the r = 0 floor, and 2.193 (0.000-7.358) for the 2012 cohort of Ontario, Quebec and British Columbia. Prince Edward Island fails the pre-fit screen and is discarded, as are the single-donor Manitoba and Alberta designs and Nova Scotia, which has no not-yet-treated donor. The extended window gives 1.612 (0.284-1.995) with only three placebo units, a nominal 50 percent interval.

These designs have essentially no power. In the clean window only ages one to three are exposed (treated share about 0.08), the null standard deviation is 0.30, and an r of 12.6 would be needed to reach the null 95th percentile; Ontario, British Columbia and Saskatchewan diverge in fiscal year 2011, before exposure. The hierarchy gives 0.99 (0.32-1.89) on the same cells. The harm-direction register lists 21 Canadian synthetic-control variants with intervals above one, from 2.018 (1.235-2.257) to 9.534 (2.887-12.952), at nominal coverage of 0.50 to 0.89 or with in-space-only uncorrected intervals (Section S60.5).

S61.4 International designs

Japan's Hokkaido routine Japanese encephalitis programme (fiscal year 2016; 46 donor prefectures) gives a gap of -0.083 (-0.389 to 0.223) with a first stage of 0.833 and r = 0.904 (0.615-1.300); its outcome is an autism composite, so it is report-only. In Sweden, the Stockholm early-adopter rotavirus design on specialist-care F84 gives r = 0.891 (0.126-3.323) and the Norrbotten hepatitis B design 1.157 (0.000-3.779), both convergence-only. The early-county PCV design is invalid: its log gap of 0.734 (0.590-0.870) exceeds the ceiling attainable for any finite r, an equal-exposure falsification in 2018-2022 reproduces the gap, every treated unit fails the fit gate, and the band-differenced null depends on Jönköping. On municipal support recipients, an outcome that is not autism, the Jönköping rotavirus design gives 4.702 (1.128-9.029) but lies outside the donor hull (pre-period RMSPE 0.128 against 0.028), its estimators disagree in sign, and its gap is present before exposure; the stacked early-adopter design, 3.766 (0.977-7.148), inherits that signal. Both are discarded. In the Norrbotten hepatitis B municipal design, 8 of 15 main-pool donors had regional infant hepatitis B programmes before 2016 but are coded unexposed, and national records contradict that block's hepatitis B exposure coding.

The Madrid, eastern German (stacked; Brandenburg, Mecklenburg-Vorpommern and Thuringia; Schleswig-Holstein; the five eastern states) and Quebec MenB designs are discarded for unbounded intervals, pre-exposure gaps or slopes, an absent first stage or, in Quebec, a slowdown that began before the May 2014 campaign. The corrected Saxony rotavirus design, 0.563 (0.000-5.425), is convergence-only. Spain's late-region design is a diagnostic without r because administrative breaks make placebos non-exchangeable. The Swedish tick-borne encephalitis negative control gives 0.015 (-0.478 to 0.566) and serves only as noise calibration. The international reproduction matches 616 of 616 checked numbers.

S61.5 Reading

Where synthetic controls converge with other designs on the same cells, they show robustness to estimator, not replication. Where they diverge, as in Canada, the divergence traces to a small first stage, few donors and pre-exposure gaps. Every harm-direction synthetic-control interval belongs to a design that is discarded, lacks nominal 95 percent coverage or is convergence-only (Section S60.5). Files: `3310_synthetic_controls_us_canada.md` and `.csv`, `3311_synthetic_controls_international.md` and `.csv`, and data rows 168-206 of `3309_jump_master_table.csv`.

S62. Generalized-claim tests and whole-pipeline power

S62.1 Estimands and the stated rule

The generalized claim states that vaccination multiplies autism risk by one factor R in exposed children in every context. Three families of tests address it. Model A fits one common individual-scale effect shared by every pooled context. Model B fits one common effect per product. Model C compares a model in which every pooled context is harmed by at least c with an all-null model. The tests reuse the primary synthesis unchanged: fifteen pooled contexts in fourteen likelihoods comprising 5,133 cells, with the same curves, cells, exposure paths and gate decisions. U.S. rotavirus 2006 stays inside the joint U.S. likelihood but untied, and the PCV-by-RotaShield joint-receipt interaction stays a free nuisance term, so the common effect is imposed only on the fifteen pooled contexts.

The rule was written at 2026-09-29T15:06:08, before any computation of these tests, and is stored as `qa/generalized_claim_20260929/3350_rule.json` with its as-written copy. A threshold is excluded at 95 percent if and only if the posterior 97.5th percentile of the target lies below it; the thresholds are RR 1.05, 1.10, 1.25 and 1.50, and P(target > threshold) is reported beside every call. EXCLUDED (robust) requires exclusion under the assumed bias floors (A1) and also with doubled floors (A3), in all fifteen leave-one-out fits (A4), without the four strict-placebo failures (A5) and without U.S. PCV (A6). EXCLUDED under the assumed floors (not robust) means A1 excludes but at least one of those checks does not. EXCLUDED only under placebo-calibrated bias SDs means A1 does not exclude but A2 does. NOT EXCLUDED means neither A1 nor A2 excludes.

The stated verification of A1 by two independent implementations failed at the specified tolerance of 1e-6 (largest difference 4.0e-05 in the log marginal likelihood) because trapezoid weights integrate the kinks of the hierarchy's linearly interpolated product functions with a small bias. Amendment 1 (`qa/generalized_claim_20260929/3350_amendment_1_quadrature.json`) changes only the numerical integration, to knot-aligned quadrature; estimands, priors, thresholds, rules and tolerance are unchanged. The amendment was written after the A and B fits and 41 of the 59 C models had been computed and inspected, and it also corrected the file names of two C model ids; every A and B row is therefore also summarised under the stated trapezoid weights. Under it the two implementations agree to 1.3e-09 on every reported quantity. The largest difference between amended and stated weights on any log-scale quantile or probability is 3.2e-05, and no exclusion flag, category or call differs.

S62.2 One common effect in every context

Each context's likelihood parameter is the common log effect plus a shared bias with standard deviation 0.15, a country-group bias (0.20 for the UK trio and the Canadian pair, 0.25 for the Sweden-Germany care bridge) and the context's own bias; all three are integrated out, and the common effect has prior N(0, log 2). Product and context heterogeneity are set to zero. Computationally this is the primary likelihood with its heterogeneity standard deviations fixed at zero.

Table S62-A. Common individual-scale effect across the pooled contexts under each stated bias setting.

Model (context bias SDs; contexts)	Common RR (95% CrI)	P(RR > 1.10)	P(RR > 1.25)	P(RR > 1.50)
A1 assumed floors, primary test (0.35 / 0.70; 15)	0.964 (0.637-1.455)	0.265	0.108	0.018
A2 placebo-calibrated, rms (0.04-0.08; 15)	0.974 (0.713-1.330)	0.221	0.058	0.003
A2u placebo-calibrated, 97.5th percentile (0.11-0.27; 15)	0.956 (0.682-1.339)	0.208	0.060	0.004
A3 doubled floors (0.70 / 1.40; 15)	0.977 (0.540-1.763)	0.348	0.207	0.078
A5 without the four strict placebo failures (0.35 / 0.70; 11)	0.950 (0.601-1.496)	0.264	0.119	0.024
A6 without U.S. PCV (0.35 / 0.70; 14)	0.955 (0.627-1.449)	0.253	0.103	0.017
A4 leave one out, 15 fits (0.35 / 0.70; 14)	medians 0.944-1.010; upper limits 1.435-1.553	0.241-0.348	0.097-0.166	0.015-0.036
Primary heterogeneous shared mean, a different estimand (0.35 / 0.70; 15)	0.957 (0.599-1.517)	0.275	0.126	0.028

Notes. Posterior medians with 95 percent credible intervals. Register and education outcomes carry context bias SD 0.35; care, composition and survey outcomes and the three core contexts carry 0.70. A2 and A2u use each context's placebo-calibrated shrunk bias SD (root mean square or 97.5th percentile). A3 doubles the context and country-group allowances; the shared bias SD 0.15 is the same in every row. P(RR > 1.05) is 0.343 under A1, 0.317 under A2 and 0.406 under A3. A2 had a minimum relevant core effective sample size of 7.9 at power 18 and was refitted at power 20, where it is 43.0, still below the floor of 50, and is reported with that flag. Source: `3350_generalized_claim_tests.csv` rows A1, A2, A2u, A3, A5, A6 and A4_*.

Under the assumed floors the common effect is 0.964 (0.637-1.455). RR 1.05, 1.10 and 1.25 are not excluded in any row. RR 1.50 is excluded under A1, A2, A2u, A5 and A6 but not under A3, whose upper limit is 1.763, or in the leave-one-out fit without Colombia varicella, whose upper limit is 1.553; the lowest leave-one-out upper limit, 1.435, comes without Scotland rotavirus. The stated category for RR 1.50 is therefore EXCLUDED under the assumed floors (not robust: fails in A3, A4), and the category for 1.05, 1.10 and 1.25 is NOT EXCLUDED. The RR 1.50 exclusion is also lost when Colombia's 2020-2021 school years are dropped from its block or absorbed as nuisance terms (Section S62.8).

S62.3 Why the common effect differs from the shared mean

The primary gives each product its own effect scattered around a shared mean, with product SD 0.1, 0.35 or 0.7 and context SD 0.1 or 0.35 averaged with equal prior weight, and reports the centre of that scatter. The common-effect model assumes no scatter, so all fifteen contexts pool fully. That assumption is the only difference: likelihoods, context allowances and shared bias are identical. The narrowing is modest, from a log-scale interval width of 0.930 to 0.826 and an upper limit of 1.517 to 1.455, because the primary already places posterior weight 0.581 on its smallest heterogeneity component (product and context SDs 0.1), where its shared mean is 0.961 (0.627-1.471), and because the bias allowances that both models keep account for most of the remaining width.

As a diagnostic rather than a test, the log marginal likelihood of A1 exceeds the primary's by 1.53, a Bayes factor of 4.64. The data therefore do not favour heterogeneous products over one common effect; a Bayes factor this small does not establish a common effect either.

S62.4 The bias allowance caps attainable precision

The shared bias moves every context together, so the data identify only the sum of the common log effect and the shared bias, never the common effect alone. With shared bias N(0, 0.15^2) and prior N(0, log 2), the common log effect retains a posterior SD of 0.1466 given that sum, however much data there are, because its conditional variance is lambda x 0.15^2 with lambda = (log 2)^2 / ((log 2)^2 + 0.15^2). Even infinitely precise data centred on RR 1.00 would leave the 97.5th percentile of a common effect at exp(1.96 x 0.1466) = 1.333. To exclude a threshold at 95 percent with no other uncertainty, the pooled data would have to be centred below RR 1.132 to exclude 1.50, below 0.935 to exclude 1.25, below 0.818 to exclude 1.10 and below 0.779 to exclude 1.05. Symmetrically, a posterior probability above 0.975 that the effect exceeds one requires the identified sum to exceed 0.301 on the log scale, RR 1.35.

The observed pooled centre under A1 is RR 0.962. At that centre the shared bias alone would give an upper limit of 1.285, and the assumed context allowances of 0.35 and 0.70 widen it to 1.455. Under the placebo-calibrated context allowances (A2) the upper limit is 1.330, already at the shared-bias floor. Removing the shared bias, a decomposition

rather than a stated test, gives 0.963 (0.705-1.310) with the assumed floors and 0.972 (0.858-1.104) with the placebo-calibrated allowances; even then RR 1.25 is not excluded under the assumed floors and RR 1.10 is not excluded under the calibrated allowances. Nothing in these data justifies dropping the shared bias, which data of this kind cannot estimate. The exclusion limits reported above therefore come from the assumed bias model, not from a shortage of observations. Section S62.10 shows how the upper limit moves with the shared-bias SD, and Section S62.9 replaces the assumed allowances with allowances calibrated on the negative controls.

S62.5 One common effect per product

Each product's contexts share one effect with prior N(0, log 2), the same bias structure and no heterogeneity; other products' contexts are removed. A single-context product is that context's own data plus its bias allowance. These are product-specific statements, not the generalized claim.

Table S62-B. Product-level common effects under the specified and placebo-calibrated bias allowances.

Product (pooled contexts)	Assumed floors: RR (95% CrI)	Placebo-calibrated: RR (95% CrI)	Thresholds excluded, calibrated	Own data, no bias allowance
Modern rotavirus (England, Scotland, Northern Ireland, Canada excluding Quebec, Quebec)	1.032 (0.611-1.734)	1.006 (0.666-1.517)	none	-
Modern rotavirus plus U.S. rotavirus 2006 (sensitivity)	1.009 (0.631-1.608)	0.976 (0.690-1.380)	1.50	-
PCV (U.S., Sweden, Germany, Hong Kong)	1.023 (0.491-2.131)	1.045 (0.752-1.457)	1.50	-
MMR (Japan, Czechia)	0.938 (0.319-2.764)	0.899 (0.344-2.315)	none	-
RotaShield (U.S.)	0.919 (0.312-2.656)	0.889 (0.373-1.925)	none	0.872 (0.361-1.832)
Maternal DTaP (Brazil)	0.874 (0.323-2.366)	0.777 (0.474-1.271)	1.50	0.766 (0.513-1.145)
Varicella (Colombia)	0.807 (0.394-1.637)	0.773 (0.474-1.219)	1.25, 1.50	0.764 (0.509-1.074)
Schedule (Tanzania)	1.066 (0.461-2.257)	1.099 (0.526-1.845)	none	1.118 (0.543-1.687)

Notes. Posterior medians with 95 percent credible intervals. No product excludes any threshold under the assumed floors. The MMR and RotaShield rows under both allowances and the calibrated modern-rotavirus-plus-2006 row remain below an effective sample size of 50 at power 20 (between 1 and 23) and are flagged; their power-20 medians and limits are within 3.7 percent of the power-18 fits. Tying U.S. rotavirus 2006 is a sensitivity only, because that context is not pool-ready. Source: `3350_generalized_claim_tests.csv` rows B_*; own-data posteriors from `hierarchy_final_p18.json` and `us_alone_diagnostics_p18.json`.

Under the assumed floors no product-level common effect excludes even RR 1.50; upper limits run from 1.64 to 2.76 for the seven products (1.61 when U.S. rotavirus 2006 is tied into modern rotavirus). A product-specific claim, for example that PCV multiplies autism risk by 1.5, is therefore far less constrained than the generalized claim, because each product has one to five contexts and each context carries a bias allowance of 0.35 or 0.70.

S62.6 Are all contexts harmful?

Model C writes each pooled context's likelihood parameter as its true effect plus the A1 bias terms. The null H0 sets every true effect to zero. The harm hypothesis H_harm(c) draws every true effect independently from a normal distribution centred at log c and truncated below at log c, with scale log 2 as the headline and 0.05, log 2 / 4, log 2 / 2 and 2 log 2 as the scale sensitivity. Stated labels are decisive against H_harm for a Bayes factor below 1/100, strong from 1/100 to 1/10, moderate from 1/10 to 1/3 and inconclusive from 1/3 to 3.

Table S62-C. Bayes factors for every pooled context harmed by at least c against the all-null model.

c	Half-normal scale	BF : H0), assumed floors	Stated label	BF(H_harm : H0), doubled floors
1.05	0.05	0.845	inconclusive	0.94
1.05	log 2 / 2	0.164	moderate against H_harm	0.477
1.05	log 2	0.00972	decisive against H_harm	0.111
1.05	2 log 2	4.8e-05	decisive against H_harm	0.00315
1.10	0.05	0.732	inconclusive	0.888

c	Half-normal scale	BF : H0), assumed floors	Stated label	BF(H_harm : H0), doubled floors
1.10	log 2 / 2	0.118	moderate against H_harm	0.412
1.10	log 2	0.00621	decisive against H_harm	0.0888
1.10	2 log 2	2.8e-05	decisive against H_harm	0.0023
1.25	0.05	0.391	inconclusive	0.686
1.25	log 2 / 2	0.0385	strong against H_harm	0.249
1.25	log 2	0.00148	decisive against H_harm	0.0434
1.25	2 log 2	5.3e-06	decisive against H_harm	8.9e-04

Notes. Marginal likelihoods are integrated numerically on the same likelihood curves and core draws as the A rows. The scale log 2 / 4 rows (0.511, 0.406 and 0.171 under the assumed floors) and the placebo-calibrated rows are in the source table. Source: `3350_generalized_claim_tests.csv` rows C_*.

The stated call treats H_harm(c) as disfavoured only if its Bayes factor is below 1/10 at every scale under the assumed floors. At the headline scale the Bayes factors are 0.00972, 0.00621 and 0.00148 for c = 1.05, 1.10 and 1.25, decisive against scattered, larger harms in every context. At the smallest scale, where H_harm(c) is nearly a uniform shift of every context by log c, they are 0.845, 0.732 and 0.391, inconclusive, because the shared bias and context allowances absorb a uniform shift. The all-harmful hypothesis is therefore scale-dependent at every c tested. This is the same shared-bias limit that prevents the common-effect test from excluding small effects. The all-null model against the heterogeneous primary has a Bayes factor of 15.1 under the assumed floors and 5.41 with doubled allowances. This is a stronger statement than the shared-mean point null of Section S40.3 (Bayes factor 2.932), which keeps product and context heterogeneity; but H0 here keeps every bias term, so a Bayes factor in its favour says that the data are well described by zero true effects plus the assumed biases, not that every effect is zero.

S62.7 Whole-pipeline power and calibration

The simulation passes data with known effects through the whole pipeline. Seventeen data-generating scenarios are applied to every pooled context on its own cells and exposure path: no harm; a common true individual RR of 1.05, 1.10, 1.25 or 1.50 in every context, including the U.S. rotavirus 2006 path; and harm at the same RR confined to modern rotavirus, PCV or MMR with every other context at 1.00. Each replicate is the block's noise-free null expectation plus the coverage-proportional mixture effect plus the block's own parametric noise from the harm-injection gate's bootstrap, including residual serial and cohort dependence and observed nuisance trends. Every block likelihood is refitted as the pipeline fits it, and the full primary hierarchy (fifteen contexts, fourteen likelihoods, U.S. rotavirus 2006 untied) and the common-effect model are recomputed. Each scenario has 400 pipeline replicates in two independent seed streams; count blocks are refitted less often and recycled, with cluster-robust Monte Carlo standard errors. The rule `qa/portfolio_power_20260929/3351_rule.json` was written before any replicate. Detection means P(target > 1) > 0.975; exclusion of c means a posterior 97.5th percentile below c.

Table S62-D. Operating characteristics of the shared mean and the common effect.

Scenario and estimand	Detection, percent (MC SE)	Excluding RR 1.25, percent	Excluding RR 1.50, percent	Median posterior median
No harm: shared mean	0.0 (0.0)	0.0	16.2	0.991
No harm: common effect	0.0 (0.0)	0.0	62.0	0.996
Common RR 1.10: shared mean	0.0 (0.0)	0.0	1.0	1.078
Common RR 1.10: common effect	0.0 (0.0)	0.0	8.5	1.088
Common RR 1.25: shared mean	0.0 (0.0)	0.0	0.0	1.202
Common RR 1.25: common effect	0.0 (0.0)	0.0	0.0	1.226
Common RR 1.50: shared mean	0.8 (0.4)	0.0	0.0	1.404
Common RR 1.50: common effect	34.2 (2.3)	0.0	0.0	1.442

Notes. 400 pipeline replicates per scenario. Under a common RR of 1.05 both estimands detect 0.0 percent. Exclusion of RR 1.05 and 1.10 is 0.0 percent in every scenario shown. Interval coverage of the true common RR is 100.0 percent for both estimands at every simulated R. Source: `3351_portfolio_power_calibration.csv`.

Neither estimand detects a common harm up to RR 1.50 with 80 percent power. Detection is 0.0 percent at a common RR of 1.05, 1.10 and 1.25 for both; at 1.50 it is 0.8 percent for the shared mean and 34.2 percent for the common effect. Under no harm, RR 1.05, 1.10 and 1.25 are never excluded, and RR 1.50 is excluded in 16.2 percent of replicates by the shared mean and 62.0 percent by the common effect; no threshold is reliably excluded. The false-positive rate under no harm is 0.0 percent for the shared mean, the common effect and the modern-rotavirus, PCV and MMR product posteriors, against a nominal one-sided 2.5 percent, and 95 percent intervals cover the true common RR in 100.0 percent of replicates at every R. True bias is zero in every scenario while the model carries its bias allowances, so these are operating characteristics in an unbiased world, not guarantees against real bias. The common-effect model is the more informative summary in both directions: under no harm its median posterior 97.5th percentile is 1.471 against 1.570 for the shared mean. The structural ceiling of S62.4 binds in the replicates: under no harm the smallest posterior 97.5th percentile is 1.359 for the shared mean and 1.301 for the common effect, and only 0.0 and 1.8 percent of replicates fall below 1.333, which is possible only when the data place the hypermean below one.

Harm concentrated in one product is diluted. With RR 1.50 confined to one product, the shared mean and the common effect detect it in 0 percent of replicates for every product, and the harmed product's own posterior detects it in 0.8 percent (modern rotavirus), 0.0 percent (PCV) and 0.0 percent (MMR); the median posterior medians of those product posteriors are 1.277, 1.097 and 1.005. A global summary nevertheless "rules out" RR 1.50 in some of these worlds: its 97.5th percentile falls below 1.50 in 2.5 percent (shared mean) and 32.8 percent (common effect) of replicates when the harm is confined to PCV, and in 11.0 and 56.2 percent when it is confined to MMR. A global exclusion therefore does not rule out a harm concentrated in one product. The block-level recovery explains the dilution: with RR 1.50 injected, the Germany PCV block's median moves only from 0.983 to 0.990, the Japan MMR block's from 0.769 to 0.895 and the Czech MMR block's from 1.026 to 1.088, whereas England rotavirus moves from 0.998 to 1.471. These product posteriors are centred on the hypermean including the shared-bias term, so the structural ceiling does not bind them.

The observed primary shared mean, 0.957 (0.599-1.517), lies at the 28th percentile of the no-harm distribution of posterior medians (null 95 percent range 0.892-1.098) and at the 0th percentile under a common RR of 1.25 or 1.50; its 97.5th percentile lies at the 25th percentile of the null upper bounds. The simulation's own observed-data computation of the common effect is 0.977 (0.659-1.446), at the 34th null percentile. It differs slightly from the reported test value, 0.964 (0.637-1.455), because the simulation uses a grid version of the joint U.S. likelihood with a point NIS exposure draw instead of the 512 bootstrap draws (that grid version reproduces the primary shared mean within 0.0012 against a tolerance of 0.005) and because its common-effect model, as specified in its rule, omits the country-group bias terms that the reported test keeps. The simulated worlds contain no systematic bias, so the 0 percent false-positive rate and 100 percent coverage describe sampling behaviour only; they are not evidence that the bias allowances are calibrated. The results are conditional on the specified exposure paths, the bias allowances, additive coverage-proportional mixture effects, independent joint receipt with no PCV-by-RotaShield interaction and each block's nuisance structure. The largest standardized difference between the two independent seed halves over all 425 rates is 1.79.

Figure S105 extends that placement to every common-effect scenario (file 3352, which refits nothing and reads the 400 per-replicate shared-mean medians that 3351 stored for each scenario). The observed shared mean, 0.957, lies above 27.75 percent of the replicate medians under no harm and above 6.5, 1.75, 0 and 0 percent of those under a common RR of 1.05, 1.10, 1.25 and 1.50, whose central 95 percent ranges are 0.931-1.143, 0.971-1.189, 1.086-1.326 and 1.267-1.554. With Colombia's lockdown-free curve (Section S62.8) the shared mean is 1.045, which lies above 83.25, 59.0, 30.5, 0.25 and 0 percent of them. These percentiles assume unbiased designs: the simulated worlds contain no systematic bias, so the percentiles describe sampling variability only and are not the formal test, which keeps the shared bias and the assumed context allowances and excludes none of RR 1.05, 1.10 and 1.25 (Section S62.2). Taken on their own terms, the observation that a common RR of 1.10 rarely produces a synthesis this low (1.75 percent of replicates) depends on Colombia's 2020-2021 school years and disappears without them (30.5 percent), whereas a common RR of 1.25 produces one this low in 0 percent of replicates with those years and 0.25 percent without them.

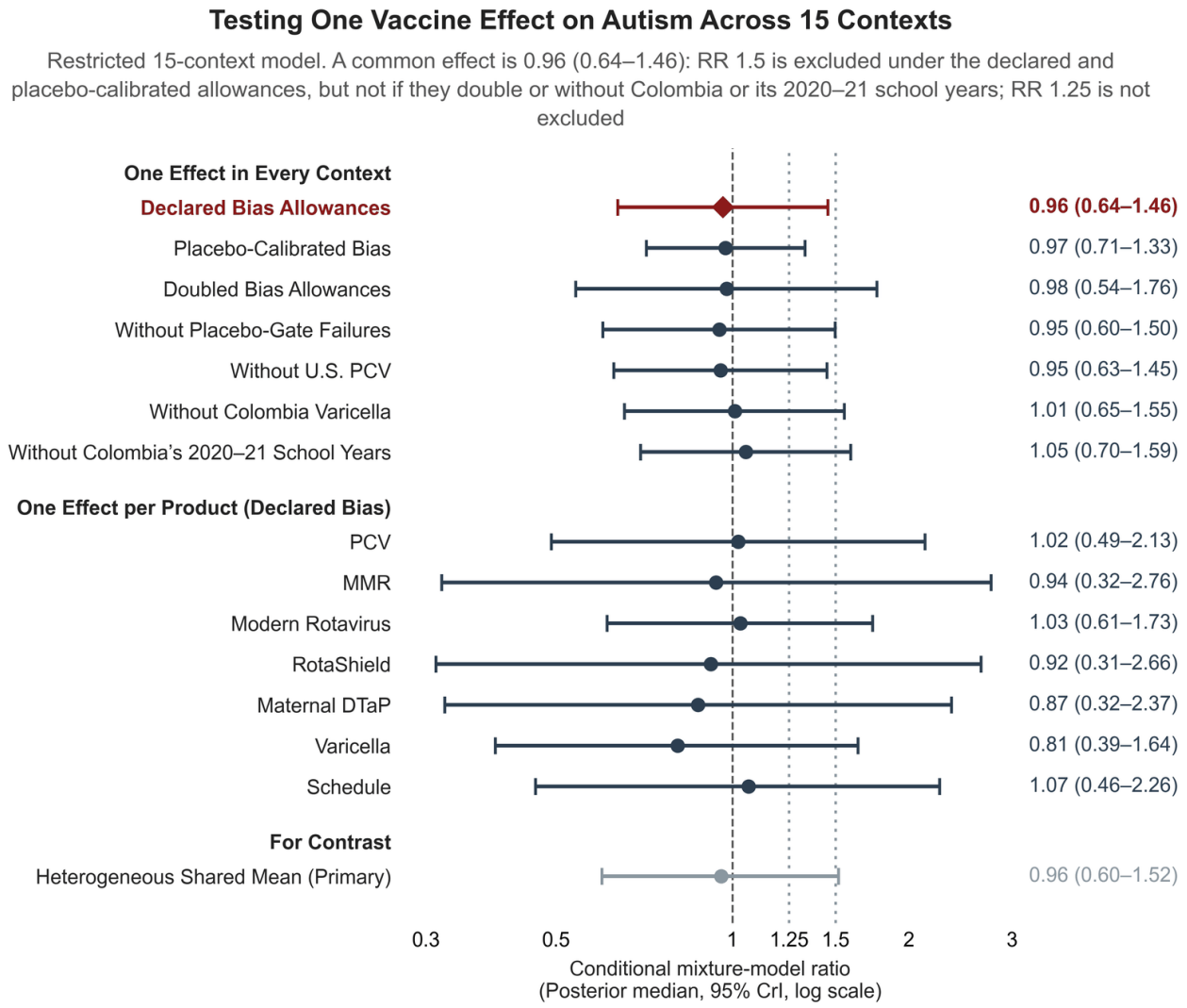


Figure S74. Generalized-claim tests. Posterior medians and 95 percent credible intervals for one common individual-scale effect across the 15 pooled contexts under the assumed bias allowances (red), under placebo-calibrated and doubled allowances, without the four contexts that fail the strict two-sided placebo gate, without U.S. PCV, without Colombia varicella, and with Colombia kept but its 2020-2021 school years removed (file 3371; Section S62.8); for one common effect per product under the assumed allowances; and, for contrast, the heterogeneous shared mean, a different estimand. Dotted lines mark RR 1.25 and 1.5. RR 1.5 is excluded under the assumed allowances but not when they are doubled, without Colombia or without its 2020-2021 school years; RR 1.25 is not excluded in any row. The shared-bias allowance caps attainable precision (Section S62.4).

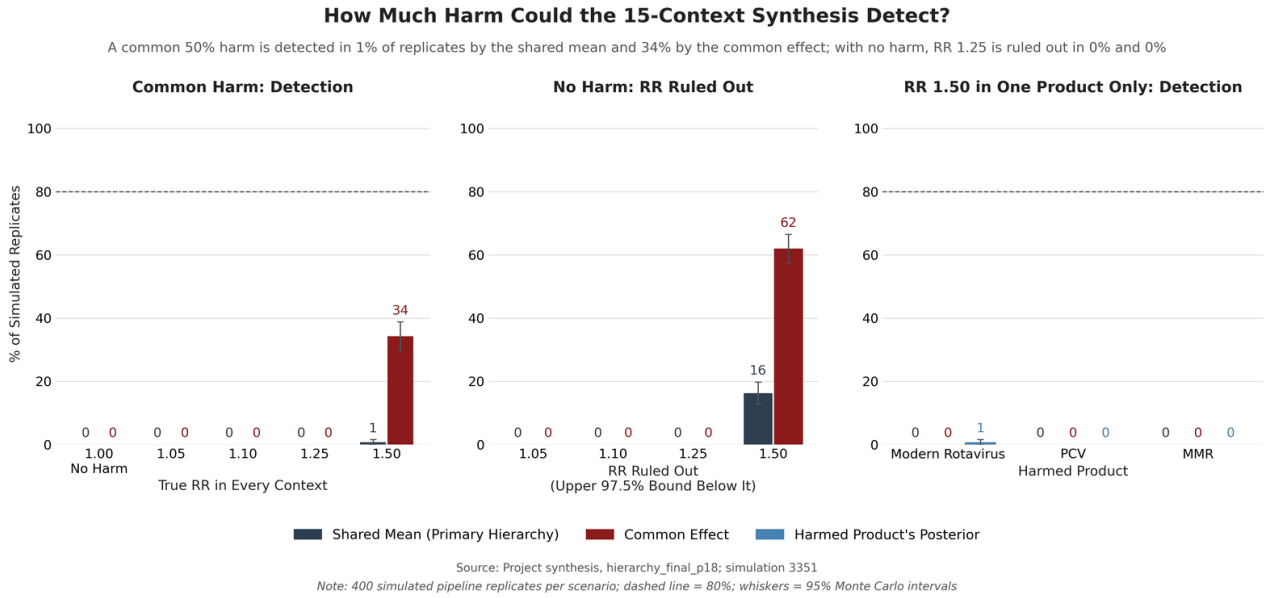


Figure S75. Whole-pipeline power and calibration. Percent of 400 simulated pipeline replicates per scenario in which the shared mean (dark) or the common effect (red) detects harm when a common true RR is imposed in every context (left), rules out a given RR when there is no harm (middle), and detects RR 1.50 confined to one product, with the harmed product's own posterior in blue (right). Detection means a posterior probability above 0.975 that the effect exceeds one. Whiskers are 95 percent Monte Carlo intervals; the dashed line marks 80 percent. True bias is zero in every scenario.

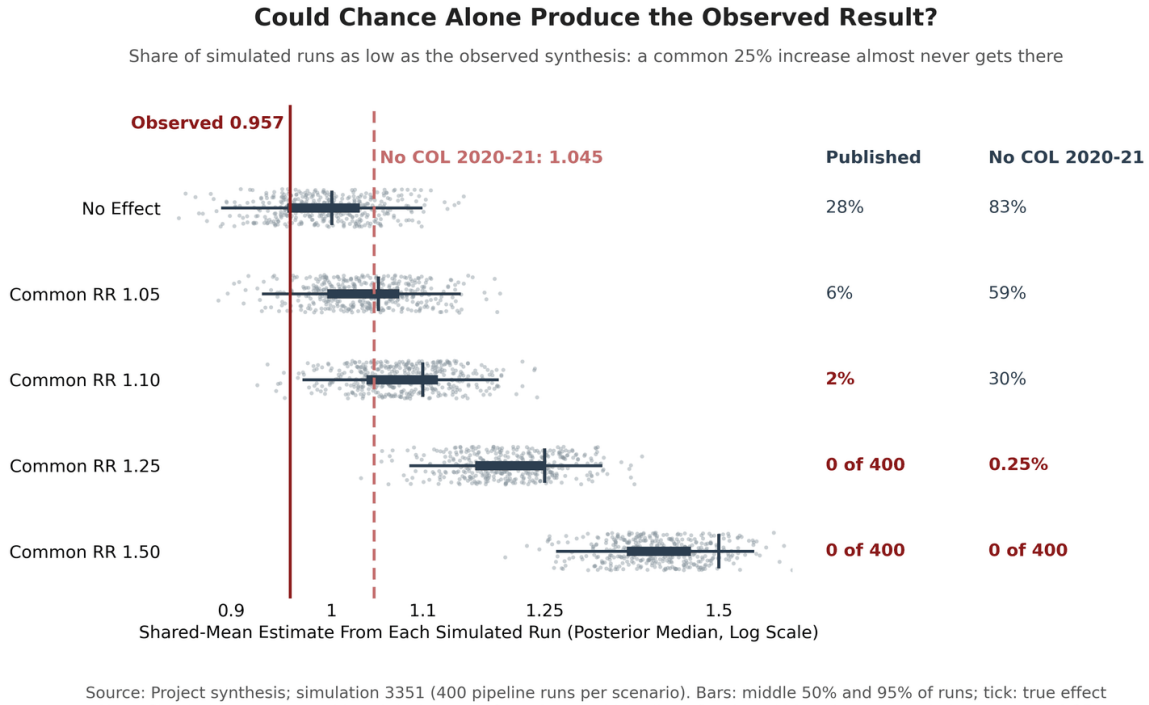


Figure S105. Observed synthesis against the whole-pipeline simulation. For no harm and for a common true RR of 1.05, 1.10, 1.25 and 1.50 in every context, dots are the shared-mean posterior medians of the 400 simulated pipeline replicates (file 3351), bars span their middle 50 and 95 percent and ticks mark the true effect. The solid line is the observed shared mean, 0.957; the dashed line is the shared mean with Colombia's 2020-2021 school years removed, 1.045. The columns give the percent of replicates below each line, rounded (0 of 400 where none is), in red where below 2.5 percent. The simulated worlds contain no systematic bias, so the percentiles describe sampling variability only and are not the formal test.

S62.8 Colombia's pandemic school years

The Colombian varicella context rests on school-census counts that include the 2020-2021 school years of the pandemic lockdowns, and the Colombian negative-control task found that the context's autism estimate moves above one without those years (Section S63.6). A sensitivity analysis (file 3371), prompted by that finding and run under a rule written before any of its own computations, carries the dependence into the synthesis. It refits the Colombian block and substitutes the refitted curve for the production curve in the primary shared mean and in the A1 common effect, with every other block, prior and setting unchanged; with the production curve both estimands reproduce their published values exactly. R1 drops the 2020 and 2021 school years (336 of 504 cells), and R2 keeps all 504 cells and absorbs level-specific 2020 and 2021 shifts as nuisance terms.

Table S62-E. The synthesis with Colombia's 2020-2021 school years removed or absorbed.

Colombian curve (cells)	Colombia own data, RR (95% CrI)	Shared mean, RR (95% CrI)	Common effect A1, RR (95% CrI)	A1 P(RR > 1.25)
Production, published (504)	0.76 (0.51-1.07)	0.957 (0.599-1.517)	0.964 (0.637-1.455)	0.108
R1: 2020-2021 removed (336)	1.32 (0.90-1.99)	1.045 (0.654-1.655)	1.054 (0.697-1.591)	0.209
R2: lockdown-year nuisance terms (504)	1.26 (0.79-2.05)	1.036 (0.648-1.642)	1.044 (0.689-1.578)	0.196
Colombia left out	-	0.999 (0.603-1.633)	1.010 (0.654-1.553)	0.166

Notes. Posterior medians with 95 percent credible intervals; Colombia own data is the block's own-data posterior under the N(0, ln 2) prior. The A1 97.5th percentile lies below 1.50 only with the production curve, so the RR 1.50 exclusion is lost in R1, in R2 and without Colombia; RR 1.05, 1.10 and 1.25 are not excluded in any row. A1 P(RR > 1.10) is 0.265, 0.420, 0.402 and 0.348 and P(RR > 1.50) 0.018, 0.047, 0.043 and 0.036 in the four rows. For the shared mean P(RR > 1.10 / 1.25 / 1.50) is 0.275 / 0.126 / 0.028 with the production curve and 0.413 / 0.220 / 0.061 in R1. Minimum relevant core effective sample sizes with the R1 curve are 124 (shared mean) and 253 (A1), both at power 18. Source: `3371_colombia_lockdown_sensitivity.md` and `.json`.

Without the 2020-2021 school years Colombia's own estimate moves from 0.76 (0.51-1.07) to 1.32 (0.90-1.99), from below one to above it, and inside the hierarchy its context posterior moves from 0.78 (0.54-1.08) to 1.29 (0.89-1.88). The shared mean moves from 0.957 to 1.045 (median +0.088, upper limit +0.138) and the common effect from 0.964 to 1.054 (+0.090 and +0.136). The move is about twice as large as leaving Colombia out (211 percent of the leave-out shift in the shared-mean median and 198 percent in the A1 median), because the lockdown-free curve replaces a downward pull with an upward pull of similar precision. Absorbing the lockdown years as nuisance terms gives nearly the same result. Both intervals still include one with room on each side, so neither estimand establishes harm, but the RR 1.50 exclusion under the assumed floors, already not robust (Section S62.2), is lost, and P(RR > 1.25) for the common effect roughly doubles, from 0.108 to 0.209. The future+3 placebo on the lockdown-free data passes, 1.14 (0.33-3.73), but with 2020-2021 removed only the 2019-to-2022 pairs remain, so it rests on 84 cells, is identical under all four kernel settings and is a weak check of timing. The primary keeps all 504 cells with the kernel's national year effects, and its gate decision is unchanged; the lockdown-free fits are reported beside it. Section S62.7 places the lockdown-free shared mean against the simulated distributions.

S62.9 Calibrating the bias allowances with the negative controls

The bias allowances used so far are assumptions: a shared bias with SD 0.15 on the log scale and context floors of 0.35 and 0.70 (Section S62.2). File 3370 replaces them with allowances measured on the negative controls of Section S63 and recomputes the common effect and the shared mean on the primary's likelihoods, cells, exposure paths and gate decisions, keeping the country-group terms (0.20 for the UK trio and the Canadian pair, 0.25 for the Sweden-Germany care bridge). Its rule, `qa/calibrated_synthesis_20260929/3370_rule.json`, was written at 2026-09-29T21:25:41, before any calibration quantity was computed. It lists every round-one control row as primary, secondary or excluded, with a reason for each exclusion (claim-relevant categories, drowning, composites, duplicates and designs that are not block designs on the individual scale), and it fixes the error model and its priors, six variants, both estimands, the quadrature and its validation, a leave-one-control-out adequacy test and the rules that choose the reported variant and estimand. The script refuses to compute or report if the rule, an amendment or a frozen input changes.

Two dated amendments followed. Amendment 1 (`3370_amendment_1.json`, 2026-09-29T22:50:11) was written after the round-one error model and its leave-one-out validation, the four reproductions and the V1 and S1 fits at power

18 had been seen (V1 common effect 0.865 (0.708-1.058)); no other variant, no power-20 fit and no per-context calibrated estimate had been computed. Because the narrow calibrated quadrature nodes leave few importance draws for the U.S. and Japanese core contexts, the rule's local effective-sample-size floor of 50 could not be met at power 20 either. The amendment therefore refits only the reported variant and V3 at power 20, reports every variant with its two independent scrambles, and flags the headline as Monte Carlo limited if the scrambles' 97.5th percentiles differ by more than 0.02 on the log scale. It also fixed how the round-two control files 3380-3385 are read: through the harmonized rows of file 3330, sorted by a written classifier that uses the rule's definitions. Amendment 2 (`3370_amendment_2.json`, 2026-09-30T03:30:30) fixes a string-matching bug in that classifier and changes no rule, although the row it restores, the Swedish all-injury union, is one the rule's own composite definition would exclude (the harmonized source roles do not mark it as a composite); the rule-literal de-nested sensitivity below applies that definition by hand. The classifier's immunization-complication pattern (t88) matched the label of the four 3381 all-injury rows, which name T80-T88 only as excluded codes, and removed them; the fix strips exclusion clauses from a label before the patterns are applied. The amendment was written after the round-one results of every variant, the round-two rows and their classification, the 3330 preview and the 3371 result had been seen, and before any round-two fit. Of the four restored rows the Swedish all-injury control, 1.056 (0.980-1.132), enters, so 45 primary controls enter instead of 44; without the fix the common effect is 0.872 (0.741-1.023) instead of 0.871 (0.741-1.020), a difference of 0.001 in the median.

The error model has two levels. Each pooled context with its own controls has a systematic error beta_c, drawn around a mean mu_b with between-context SD sigma_b, and its control outcomes scatter around beta_c with SD omega. Each control enters as its estimate with the N(0, (log 2)^2) prior removed, and controls whose posterior log-SD exceeds 0.8 log 2 are dropped as prior-dominated. Across rounds one and two, 45 same-cell controls in 9 pooled contexts enter: Brazil maternal DTaP, Canada rotavirus excluding Quebec, Colombia varicella, England rotavirus, Hong Kong PCV, Sweden PCV, Tanzania, U.S. PCV and U.S. RotaShield (counts in Table S62-G). They give sigma_b = 0.20 (0.03-0.47), omega = 0.02 (0.00-0.09) and mu_b = +0.042 (-0.155 to +0.217), with posterior SD 0.091. The reported variant, V1, removes from each context the autism bias its controls predict, as a mean shift and a variance, plus the placebo timing bias. A context without controls (Japan MMR, Germany PCV, Scotland, Northern Ireland and Quebec rotavirus, and Czechia MMR) carries the population predictive, with a bias SD of 0.26 in place of the assumed 0.35 or 0.70. The shared component becomes the posterior uncertainty of mu_b, SD 0.091, in place of the fixed 0.15. It loads fully only on the six contexts without controls; a context with its own controls has its bias measured, and its loading on mu_b runs from 0.04 (Sweden PCV) to 0.80 (U.S. RotaShield).

Leave-one-control-out validation, the stated adequacy test, recalibrates each of the 45 controls with the error model fitted without it: 0.933 of the calibrated intervals cover RR 1, against 0.867 uncalibrated and a threshold of 0.85, so V1 is the reported variant. Because the uncalibrated intervals already clear the threshold, the gate tests only that calibration does not worsen coverage; calibration repairs three of the six uncalibrated misses (U.S. PCV hearing, Brazil falls, Sweden S42) and leaves three. Under V1 the log Bayes factor of a common effect against the heterogeneous hierarchy is 2.43, which is at least 0, so the rule takes the common effect. **The calibrated common effect is 0.871 (0.741-1.020)**, with P(RR > 1.05) = 0.010, P(RR > 1.10) = 0.002 and P(RR > 1.25) < 0.001; RR 1.05, 1.10, 1.25 and 1.50 are all excluded at 95 percent. The calibrated shared mean is 0.877 (0.696-1.102), which excludes RR 1.25 and 1.50. Under the assumed allowances the same estimands are 0.964 (0.637-1.455) (A1) and 0.957 (0.599-1.517), so calibration shrinks the log-scale interval width from 0.826 to 0.319 for the common effect and from 0.930 to 0.459 for the shared mean.

The interval narrows for two reasons. The assumed shared bias loads on every context and no data can learn it (Section S62.4); on its own it caps the common effect's upper limit near 1.33. And calibration pulls the context estimates together, because the controls lean above one where autism leans above one and below one where it leans below: U.S. PCV moves from 1.165 (0.970-1.393) on its own data to 0.886 (0.630-1.238), Hong Kong PCV from 0.892 to 0.782, and Brazil maternal DTaP up from 0.766 to 0.940 (Table S62-G). The U.S. PCV move, the largest, reads that context's control lean (bias mean +0.285, SD 0.153) as drift shared with autism, which its stated MIXED verdict (Section S63.2) could not establish. Without any mean shift (V2) the common effect is 0.946 (0.806-1.111): the exclusion of RR 1.25 does not depend on that reading, the exclusions of 1.05 and 1.10 do. Contexts without controls move by the mean bias mu_b and widen by its predictive SD, and every calibrated context interval includes one. Placebo-only calibration (V3) gives 0.972 (0.858-1.104) but is not a defensible summary: placebo breakpoints measure only timing bias (tau 0.04-0.08), whereas the controls show a non-timing bias between contexts of sigma_b = 0.20, which V3 ignores.

Table S62-F. The calibrated synthesis under each variant and sensitivity.

Model	Status	Shared mean, RR (95% CrI)	Common effect, RR (95% CrI)	Common effect P(RR > 1.10)	Common effect P(RR > 1.25)	Thresholds excluded by the common effect
Assumed allowances (primary; A1)	reference	0.957 (0.599-1.517)	0.964 (0.637-1.455)	0.265	0.108	1.50
V1 NC and placebo (reported)	stated	0.877 (0.696-1.102)	0.871 (0.741-1.020)	0.002	<0.001	1.05, 1.10, 1.25, 1.50
V2 variance only	stated	0.934 (0.731-1.176)	0.946 (0.806-1.111)	0.033	<0.001	1.25, 1.50
V3 placebo only	stated	0.945 (0.759-1.157)	0.972 (0.858-1.104)	0.029	<0.001	1.25, 1.50
V4 NC only	stated	0.875 (0.699-1.095)	0.868 (0.744-1.010)	0.001	<0.001	1.05, 1.10, 1.25, 1.50
V5 with secondary NCs	stated	0.881 (0.697-1.115)	0.875 (0.748-1.022)	0.002	<0.001	1.05, 1.10, 1.25, 1.50
S1 context mean only	stated	0.877 (0.697-1.100)	0.871 (0.742-1.019)	0.002	<0.001	1.05, 1.10, 1.25, 1.50
V1 without the amendment-2 fix	amendment 2	-	0.872 (0.741-1.023)	0.002	<0.001	1.05, 1.10, 1.25, 1.50
V1, round-two categories de-nested	post hoc	-	0.876 (0.740-1.033)	0.004	<0.001	1.05, 1.10, 1.25, 1.50
V1, Colombia without 2020-2021 (3371 R1)	post hoc	-	0.943 (0.802-1.105)	0.028	<0.001	1.25, 1.50
V1 construction on the five C_INJ controls	post hoc	-	1.064 (0.711-1.564)	0.428	0.207	none

Notes. Posterior medians with 95 percent credible intervals; a threshold is excluded when the 97.5th percentile lies below it. V1 removes each context's predicted bias from its negative controls (mean and variance) plus its placebo timing bias, with the posterior uncertainty of mu_b as the shared component; V2 removes every mean shift; V3 uses placebo timing only, with no shared term (the decomposition of Section S62.4); V4 drops the timing term; V5 adds the secondary controls to the error model; S1 keeps only each context's mean bias and is never the answer. The V1 and V3 common effects are at power 20 and the other rows at power 18; each row's two independent scrambles are in the source. The de-nested row drops 35 nested round-two categories by the rule's own composite and nested definitions, leaving 38 controls; the Colombian row substitutes the 3371 R1 autism curve, with the Colombian controls still fitted on all 504 cells; the C_INJ row fits the 3370 error model to the five injury controls of the 3330 preview. Source: `3370_calibrated_synthesis.md` and `.json`.

Round two moved the answer little. Round one alone, 19 controls in 6 contexts (sigma_b 0.256, omega 0.130, mu_b +0.066), gave a common effect of 0.864 (0.708-1.057) and a shared mean of 0.871 (0.670-1.133); rounds one and two, 45 controls in 9 contexts, give 0.871 (0.741-1.020) and 0.877 (0.696-1.102). The largest round-two shift in any variant is in V5, whose common effect moves from 0.910 (0.756-1.108) to 0.875 (0.748-1.022). Three post hoc checks are reported beside the answer and are never the answer. The specified round-two classifier drops a union only when its source calls it a composite, so nested round-two categories (for example Swedish S00-S09 head injuries beside the S06 and S02 codes, and STATS19 all casualties beside killed-or-seriously-injured and pedestrian casualties) enter as independent controls; applying the rule's own composite and nested definitions by hand (the rule-literal version, computed post hoc), before any round-two fit, leaves 38 controls (sigma_b 0.158, omega 0.062, mu_b +0.054) and a common effect of 0.876 (0.740-1.033). Colombia's lockdown-free curve (Section S62.8) gives 0.943 (0.802-1.105). The five injury controls of the 3330 preview (C_INJ; Section S63.13) under the 3370 error model give sigma_b = 0.55 (0.06-1.13), a predictive bias SD of 0.75 for a context without controls and a common effect of 1.064 (0.711-1.564); with one control per context, sigma_b and omega are not separately identified there.

The reading these results support is narrower than the headline. Common increases of 25 percent or more are excluded by both estimands in every stated calibrated variant, and by the common effect without Colombia's 2020-2021 school years. An increase of 5 percent is excluded only by the common effect in the variants that keep the controls' mean shifts (V1, V4, V5 and S1), and one of 10 percent by those and by the V4 and S1 shared means; neither is excluded when the mean shifts are removed (V2), under placebo-only calibration (V3), by the reported V1 shared mean, or without Colombia's 2020-2021 school years, where P(RR > 1.10) is 0.028. The answer also depends on using the full set of controls: with only the five C_INJ controls the common effect is 1.064 (0.711-1.564) and excludes nothing.

The calibration rests on exchangeability. Autism's design bias in each context is assumed to behave like the controls' biases in that context, contexts are assumed exchangeable with each other, and an independent timing bias is taken from the placebo breakpoints. The controls calibrate what they share with autism, such as reporting systems, denominators, exposure reconstruction and trends. A systematic error specific to autism series, for example diagnostic-practice or awareness trends that move autism identification but not injury, sensory or physical-disability counts, is not calibrated away. Only 9 of the 15 pooled contexts have controls that enter, so the other six are calibrated only as well as contexts are exchangeable, and sigma_b, estimated from 9 contexts, is correspondingly uncertain. Controls enter through a Gaussian approximation of their posteriors, overlapping Brazilian categories are treated as distinct outcomes, and the U.S. PCV and RotaShield controls come from one joint fit but are treated as separate contexts.

Numerically, the narrow calibrated nodes leave the core integrals Monte Carlo limited. The minimum relevant core effective sample size of the reported estimate is 16.4, below the rule's floor of 50, and it is flagged as the rule prescribes. Its two independent scrambles give 0.871 (0.740-1.022) and 0.871 (0.744-1.017); their 97.5th percentiles differ by 0.0048 on the log scale, within the amendment's 0.02 threshold, so the headline is not Monte Carlo limited. That threshold was specified after the round-one power-18 scrambles, which differed by 0.011, had been seen. Applied to the shared mean, the same threshold would flag V1 (0.021), V3 (0.025), V4 (0.023) and S1 (0.021) as Monte Carlo limited; their scrambles agree on every exclusion call except RR 1.10, where for V1, V4 and S1 one scramble's 97.5th percentile lies below 1.10 and the other's above it. With zero shifts and the shared SD of 0.15, the same machinery reproduces the primary, 0.957 (0.599-1.517), with components within 9.1e-13 on the log scale, and A1, 0.964 (0.637-1.455), with likelihood values within 1.8e-12. Figure S108 shows the variants and the per-context shifts.

Table S62-G. Each pooled context's own-data autism estimate before and after calibration (V1).

Context	Controls entering	Own data, uncalibrated	Own data, V1 calibrated	Bias removed, log: mean (SD)	Loading on mu_b
U.S. PCV	4	1.165 (0.970-1.393)	0.886 (0.630-1.238)	+0.285 (0.153)	0.33
U.S. RotaShield	3	0.872 (0.361-1.832)	0.826 (0.339-1.849)	+0.101 (0.227)	0.80
Japan MMR	0	0.849 (0.279-2.322)	0.840 (0.274-2.367)	+0.038 (0.271)	1.00
Sweden PCV	16	1.012 (0.786-1.555)	0.963 (0.716-1.500)	+0.056 (0.077)	0.04
Germany PCV	0	1.034 (0.267-3.962)	1.033 (0.267-3.957)	+0.038 (0.270)	1.00
Hong Kong PCV	2	0.892 (0.657-1.165)	0.782 (0.529-1.128)	+0.145 (0.137)	0.36
England rotavirus	3	0.943 (0.864-1.021)	0.923 (0.776-1.095)	+0.022 (0.077)	0.14
Scotland rotavirus	0	1.551 (0.376-5.315)	1.525 (0.378-5.177)	+0.038 (0.272)	1.00
Northern Ireland rotavirus	0	1.012 (0.878-1.155)	0.972 (0.600-1.609)	+0.038 (0.264)	1.00
Canada rotavirus (excluding Quebec)	1	0.992 (0.320-1.894)	1.055 (0.326-2.235)	-0.149 (0.201)	0.55
Quebec rotavirus	0	1.067 (0.741-1.476)	1.021 (0.579-1.791)	+0.038 (0.267)	1.00
Czechia MMR	0	1.035 (0.323-3.660)	1.032 (0.317-3.659)	+0.038 (0.271)	1.00
Brazil maternal DTaP	9	0.766 (0.513-1.145)	0.940 (0.565-1.580)	-0.221 (0.186)	0.44
Colombia varicella	6	0.764 (0.509-1.074)	0.707 (0.454-1.054)	+0.091 (0.115)	0.21
Tanzania schedule	1	1.118 (0.543-1.687)	1.082 (0.522-1.791)	+0.021 (0.153)	0.44

Notes. Own-data estimates use each context's likelihood alone, with prior N(0, (log 2)^2) and no pooling. The bias column is the V1 predictive bias, mean and SD on the log scale, that calibration removes; a positive mean moves the estimate down. Contexts without controls carry the population predictive. The loading is the weight of the shared component mu_b in the context's bias; it is 1 for the six contexts without controls. Every calibrated interval includes one. Source: `3370_calibrated_synthesis.md`, per-context and error-model tables.

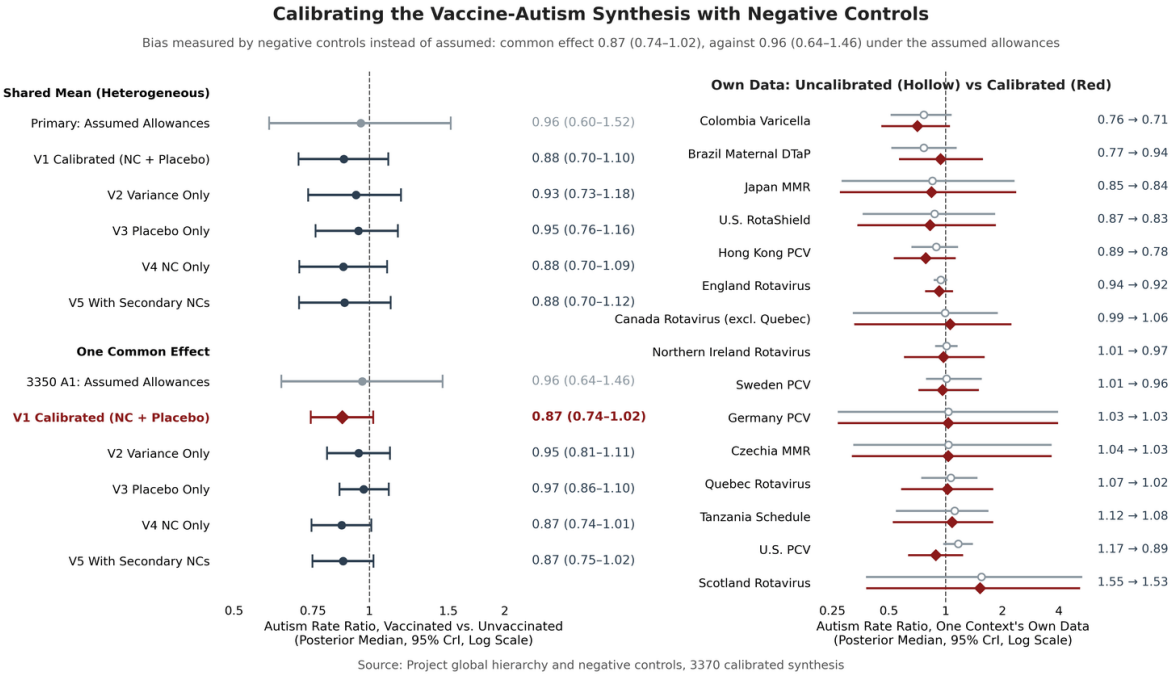


Figure S108. Calibrating the synthesis with the negative controls (file 3370). Left: posterior medians and 95 percent credible intervals of the heterogeneous shared mean (top) and the common effect (bottom) under the assumed bias allowances (grey) and the stated calibrated variants: V1 calibrates with the negative controls and the placebo breakpoints, V2 keeps only the variances, V3 only the placebo timing, V4 only the controls, and V5 adds the secondary controls. The restricted 15-context V1 common effect is 0.87 (0.74-1.02) (red). Right: each pooled context's own-data autism estimate, uncalibrated (hollow grey) and V1-calibrated (red diamond), with the medians before and after. The calibration assumes that autism's design bias in each context is exchangeable with the controls' biases; an error specific to autism series is not calibrated away (Section S62.9).

S62.10 Bias-signal contexts and the shared-bias allowance

File 3372 runs two sensitivities that keep the assumed allowances of Section S62.2. Its rule, `qa/bias_sensitivity_20260930/3372_rule.json`, was written at 2026-09-30T06:06:03-05:00, before any 3372 computation (the first fit started at 06:10:15). Neither sensitivity changes the primary, and the no-drop and single-drop fits reproduce the published primary, A1 and leave-one-out values exactly.

The first drops, jointly, the three pooled contexts whose designs show a bias signal on a primary negative control (Section S63.13): U.S. PCV (IDEA hearing impairment 1.41 (1.05-1.91)), Brazil maternal DTaP (SUS injury admissions 0.51 (0.29-0.83) and SIM unintentional injury deaths 0.36 (0.15-0.79)) and Hong Kong PCV (police road-traffic casualties 0.66 (0.48-0.92)). The fourth such context, U.S. rotavirus 2006, is already untied from the shared mean. U.S. PCV is untied rather than removed, as in the published leave-one-out fit, so its data stay in the joint U.S. likelihood. Without the three, the primary shared mean is 0.966 (0.583-1.585) and the A1 common effect 0.971 (0.630-1.491). The medians move by +0.008 and +0.007 and the upper limits by +0.068 and +0.036; P(RR > 1.25) rises from 0.126 to 0.150 and from 0.108 to 0.124; and no exclusion call changes, so A1 still excludes RR 1.50 and nothing smaller. The three contexts pull in opposite directions: alone, dropping U.S. PCV moves the A1 median by -0.009, dropping Brazil by +0.011 and dropping Hong Kong by +0.004. With Colombia's lockdown-free curve as well (Section S62.8) the estimates are 1.072 (0.648-1.755) and 1.078 (0.700-1.655), and the A1 exclusion of RR 1.50 is lost, as it is with the Colombian curve alone, which accounts for 77 to 79 percent of that median shift. Every one of these intervals includes one with room on both sides. The drop does not show that the remaining contexts are free of design bias, because several have only uninformative controls or none.

The second varies the shared-bias SD with every context and country-group allowance unchanged, by post-processing the no-drop likelihood values (the construction of 3350's decomposition). The central estimate barely moves (A1 medians 0.963-0.968), but the upper limit depends strongly on the SD (Table S62-H and Figure S109): the A1 upper limit is 1.31 with no shared bias, 1.46 at the primary's 0.15 and 1.76 at 0.30, and the shared bias's share of the posterior variance of the common effect rises from 0 percent at SD 0 to 48 percent at 0.15 and 81 percent at 0.30. On the specified grid a common RR of 1.50 is excluded by A1 at SDs of 0.15 or less and by the primary shared mean at 0.10 or

less. RR 1.25 is excluded at no SD, not even with no shared bias: under the assumed context floors the A1 upper limit is then 1.31, the value of Section S62.4, so the context allowances alone keep 1.25 inside the interval.

Table S62-H. The A1 common effect and the primary shared mean under different shared-bias SDs.

Shared-bias SD	NC support (3372 rule)	Analytic ceiling	A1 common effect, RR (95% CrI)	A1 P(RR > 1.25)	Primary shared mean, upper limit	RR 1.50 excluded, A1 / primary
0.00	not supported	1.000	0.963 (0.705-1.310)	0.048	1.393	yes / yes
0.05	not supported	1.103	0.963 (0.695-1.328)	0.056	1.407	yes / yes
0.10	not supported	1.214	0.963 (0.671-1.379)	0.078	1.450	yes / yes
0.15 (primary)	consistent	1.333	0.964 (0.637-1.455)	0.108	1.517	yes / no
0.20	consistent	1.457	0.965 (0.601-1.548)	0.142	1.601	no / no
0.30	more conservative	1.715	0.968 (0.531-1.763)	0.202	1.800	no / no

Notes. Only the shared-bias SD differs between rows; the context-bias floors (0.35 / 0.70) and country-group terms (0.20 / 0.25) are unchanged. The analytic ceiling is the 97.5th percentile of a common effect given infinitely precise data centred on RR 1, exp(1.96 sqrt(lambda) s) with lambda = (log 2)^2 / ((log 2)^2 + s^2). NC support is the classification specified in the 3372 rule before any result (see the text). The SD 0 row equals the decomposition of Section S62.4, and the 0.15 row equals the published A1 and primary. RR 1.05, 1.10 and 1.25 are excluded in no row by either estimand. The file also reports the three SDs named by the negative-control summaries, 0.112, 0.146 and 0.269, with A1 upper limits of 1.396, 1.448 and 1.693. Source: `3372_bias_context_and_shared_bias_sensitivity.md`.

What the controls support was set by the 3372 rule before any result. The three negative-control summaries it names estimate a shared bias with an uncertainty of 0.112 (se(mu) of the 3330 preview set S7, 12 controls), 0.146 (the posterior SD of mu_b in 3370) and 0.269 (se(mu) of the preview set C_INJ, 5 controls). A shared SD from 0.112 to 0.269 is therefore classed CONSISTENT with them, a smaller one NOT SUPPORTED and a larger one MORE CONSERVATIVE. Within that range the A1 upper limit runs from 1.40 to 1.69 and P(RR > 1.25) from 0.084 to 0.185, and A1's exclusion of RR 1.50 holds at 0.112, 0.146 and 0.15 and fails at 0.20 and 0.269. None of the summaries distinguishes the mean bias from zero. The range is a floor, not an estimate: a shared error specific to autism series, such as diagnostic-practice or awareness trends that move autism but not injuries, is invisible to every control, so the controls argue against shrinking the shared allowance below that floor, not for any particular larger value. 3372 read the round-one 3370 report (19 controls in 6 contexts). With rounds one and two, 3370's posterior SD of mu_b is 0.091 (Section S62.9); substituted into the 3372 rule, which 3372 did not recompute, it would lower the floor to 0.091 and make an SD of 0.10 consistent, at which the A1 upper limit is 1.38 and RR 1.25 is still not excluded.

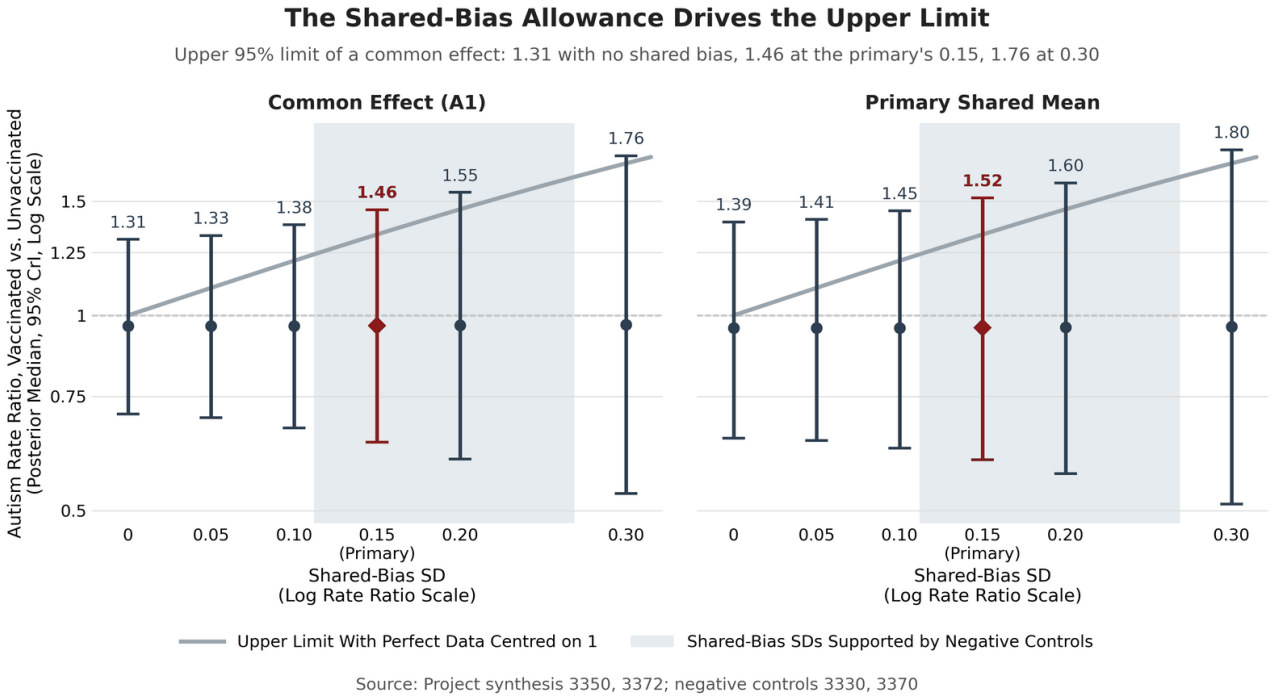


Figure S109. The shared-bias allowance and the upper limit (file 3372). Posterior medians and 95 percent credible intervals of the A1 common effect (left) and the primary shared mean (right) at shared-bias SDs from 0 to 0.30 on the log rate-ratio scale, with every context and country-group allowance unchanged; the primary's 0.15 is in red and the upper limits are printed. The grey line is the analytic ceiling, the upper limit that infinitely precise data centred on RR 1 would leave. The shaded band, 0.112 to 0.269, is the range of shared-bias SDs consistent with the three negative-control summaries named in the 3372 rule, which read the round-one 3370 report; because an error specific to autism series is invisible to the controls, the band is a floor for the shared SD, not an estimate. The medians barely move, and RR 1.25 lies inside every interval (Section S62.10).

S62.11 What these tests cannot establish

The tests do not address product-specific or subgroup effects outside the pooled contexts, including susceptible subgroups whose effect is diluted in population jumps; acute or short-latency effects and regression events not captured by birth-cohort autism counts at the observed ages; real effects that coincide with a systematic error common to every context; outcomes other than autism; or contexts that failed the gates. They inherit every modelling choice of the primary: the jump-to-individual-scale mixture conversion, exposure reconstructions, trend kernels, gate decisions made during the project rather than externally preregistered, the retention of pandemic-era school years in the Colombian block (Section S62.8), and bias allowances that are assumptions, not estimates. Section S63.13 reports what the negative controls say about those allowances. Section S62.9 replaces the assumed allowances with allowances calibrated on those controls under an exchangeability assumption, and the calibrated common effect is 0.871 (0.741-1.020). Files: `3350_generalized_claim_tests.md`, `.csv` and `.json`; `3351_portfolio_power_calibration.md`, `.csv` and `.json`; `3352_observed_vs_simulated.csv`; `3371_colombia_lockdown_sensitivity.md`, `.csv` and `.json`; `3370_calibrated_synthesis.md`, `.csv` and `.json`; `3372_bias_context_and_shared_bias_sensitivity.md`, `.csv` and `.json`; scripts `07_analysis_code/3350_generalized_claim_tests.py`, `07_analysis_code/3351_portfolio_power_calibration.py` and `07_analysis_code/3370_calibrated_synthesis.py` with their tests, `07_analysis_code/3352_observed_vs_simulated.py`, `07_analysis_code/3371_colombia_lockdown_sensitivity.py` and `07_analysis_code/3372_bias_context_and_shared_bias_sensitivity.py`; the 3371 rule `qa/colombia_lockdown_20260930/3371_rule.json`; the 3370 rule `qa/calibrated_synthesis_20260929/3370_rule.json` with `3370_amendment_1.json`, `3370_amendment_2.json` and the control manifest `3370_nc_manifest.json` in the same folder; and the 3372 rule `qa/bias_sensitivity_20260930/3372_rule.json`.

S62.12 Posterior probabilities of harm, null and benefit

File 3373 converts four posteriors of Table S62-F into the probabilities of three regions of the individual-scale rate ratio: benefit (RR < 0.95), a practical-null band of plus or minus 5 percent (0.95 <= RR <= 1.05) and harm (RR > 1.05), with the larger harms RR > 1.10 and RR > 1.25 beside them (Table S62-I). Under the calibrated common effect (V1) most of the posterior lies below 0.95 (0.859), but this reflects the subtraction of the controls' measured bias, chiefly the U.S. PCV controls' lean (Section S62.9), under the exchangeability assumption; it is not evidence of protection, and without the mean shifts (V2) P(RR < 0.95) is 0.518. At even prior odds the point-null Bayes factors give posterior null probabilities of 0.746 for the shared-mean null of Section S40.3 (Bayes factor 2.932) and 0.938 for the all-null model of Section S62.6 (15.06), or 0.844 with the bias floors doubled (5.41); both nulls keep every bias term. Nothing was refitted: before computing any probability, the script reproduced each of the four published estimates exactly from the saved fits (largest absolute difference 1.1e-13).

Table S62-I. Posterior probabilities of benefit, practical null and harm for four estimands.

Estimand	Median RR (95% CrI)	P(RR < 0.95)	P(0.95 <= RR <= 1.05)	P(RR > 1.05)	P(RR > 1.10)	P(RR > 1.25)
Shared mean, assumed allowances (primary)	0.957 (0.599-1.517)	0.487	0.168	0.345	0.275	0.126
Common effect, assumed allowances (A1)	0.964 (0.637-1.455)	0.472	0.185	0.343	0.265	0.108
Common effect, calibrated (V1, reported)	0.871 (0.741-1.020)	0.859	0.131	0.010	0.002	<0.001
Common effect, calibrated without mean shifts (V2)	0.946 (0.806-1.111)	0.518	0.379	0.102	0.033	<0.001

Notes. The quantity is the individual-scale rate ratio of a vaccinated versus an unvaccinated child. The first three probability columns partition the posterior; P(RR > 1.25) is 4.6e-06 under V1 and 3.7e-04 under V2. The band is symmetric on the ratio scale, unlike the reciprocal margin of Section S40, so the primary's probabilities differ slightly from those reported there. Source: `3373_posterior_regions.md` and `.json`.

S63. Negative-control outcomes

S63.1 Design and stated rules

A negative-control outcome replaces autism with an outcome that vaccination cannot plausibly cause and keeps everything else: the design's cells, exposure path, denominators, trend kernels, priors and likelihood code, imported read-only with only the outcome swapped. Each task wrote its rule file before its first fit and refuses to report if the rule changes; all tasks but the earliest, the U.S. special-education task (file 3326), whose fits predate hash recording and rest on file times alone, also recorded the rule hash in every fit receipt. Injury tasks use a common three-way classification. A control is UNINFORMATIVE if the log width of its 95 percent interval exceeds 2.5 times that of the matching autism estimate on the same cells; several tasks add criteria for zero cells, thin counts or prior-dominated posteriors. Otherwise it is CONSISTENT if its interval includes one and a BIAS SIGNAL if it does not, with the direction reported. Each system then has a stated reading; for example, the Brazilian rule reads DESIGN BIAS DETECTED if the all-injury composite is a bias signal, or if at least two informative components are bias signals in the same direction. Future-exposure placebos, autism-relative-to-control log-ratio fits (DDD), harm-injection and plasmode gates, kernel and prior stability and alternative cell rules are reported beside the class and never change it.

A bias signal means the design returns an "effect" where none can exist, so its estimates on those cells carry non-specific bias and are not interpretable as vaccine effects; the rules do not say how much of an autism lean that bias explains. CONSISTENT means no bias was detected at the control's precision; it does not validate the autism estimate. UNINFORMATIVE means no conclusion. Other conditions are not automatically negative controls: intellectual disability, speech, learning, emotional and other health impairments are claim-relevant and never enter a verdict.

A second round of tasks (files 3380-3385) fitted series acquired after the first round: English hospital admissions and road casualties, Scottish admissions and road casualties and Northern Irish road casualties for the UK designs; Swedish,

German and Czech care-register injuries for the pooled European designs; Brazilian SIM injury deaths; U.S. child injury deaths from the NCHS multiple-cause microdata; asthma from the Canadian Chronic Disease Surveillance System; and injury series for the Colombian, Hong Kong, Japanese and Taiwanese designs. Each round-two task used the same three-way classification with the 2.5-fold width criterion (with the round-one thinness and prior-dominance criteria where its round-one counterpart had them), wrote its rule file before its first fit and recorded the rule hash in every fit record. The Brazilian rule was written before any death count was seen and the U.S. rule before the death files were parsed. The other round-two rules were written after their data had been downloaded (for file 3381, all but the Swedish external-cause series reported beside its readings, which finished downloading after the rule), and the counts of file 3385, including the Yokohama counts, and of the 3381 reading designs were built before their rules; 3385 adopted the round-one Japanese criteria unchanged and 3381 the round-one German criteria.

S63.2 U.S. special-education categories

The U.S. core PCV context leans harmful on its own data, 1.17 (0.97-1.39) with P(r > 1) = 0.948. The first negative-control task refits the same joint U.S. likelihood on the same state-by-cohort cells, with autism replaced by other special-education categories treated as candidate controls. Its rule differs from the injury rule. A control is UNINFORMATIVE if its interval is more than twice as wide as autism's on the log scale (or its effective sample size stays below 100); otherwise SIMILAR if its median exceeds one, P(r > 1) is at least 0.80 and its interval overlaps autism's, NULL-CENTRED if its interval includes one and its median lies in [0.95, 1.05], and BELOW or OTHER otherwise. The verdict is NON-SPECIFIC if at least two informative controls are SIMILAR and outnumber NULL-CENTRED ones, AUTISM-SPECIFIC if at least two are NULL-CENTRED, none is SIMILAR and autism's P(r > 1) is at least 0.80, and MIXED otherwise. The script reproduces the core U.S.-alone posterior exactly with the autism outcome.

Table S63-A. U.S. core PCV posterior with autism replaced by other special-education categories.

Outcome (role; cells)	PCV r (95% CrI)	P(r > 1)	Class; future+3 placebo	Rotavirus 2006 r (95% CrI)
Autism (reference; 1,469)	1.17 (0.97-1.39)	0.948	reference; PASS	0.95 (0.76-1.19)
Traumatic brain injury (negative control; 472)	1.65 (0.77-4.29)	0.892	UNINFORMATIVE, width 4.75x; PASS	2.37 (1.09-4.58)
Orthopedic impairment (negative control; 1,261)	1.43 (0.91-2.26)	0.937	UNINFORMATIVE, width 2.52x; PASS	1.01 (0.70-1.46)
Hearing impairment (negative control; 1,381)	1.41 (1.05-1.91)	0.990	SIMILAR, width 1.65x; PASS	1.41 (1.07-1.82)
Visual impairment (negative control; 1,073)	1.45 (0.74-2.72)	0.855	UNINFORMATIVE, width 3.61x; PASS	1.16 (0.78-1.71)
Intellectual disability (claim-relevant; 1,117)	1.18 (0.94-1.49)	0.925	not a control; FAIL	1.42 (0.96-2.05)
Speech or language impairment (claim-relevant; 1,459)	1.09 (0.95-1.24)	0.898	not a control; FAIL	1.11 (0.94-1.31)
Specific learning disability (claim-relevant; 845)	1.13 (0.95-1.33)	0.918	not a control; FAIL	0.89 (0.58-1.36)
Emotional disturbance (claim-relevant; 809)	1.02 (0.79-1.30)	0.561	not a control; PASS	0.47 (0.20-0.86)
Other health impairment (claim-relevant; 1,364)	1.40 (1.12-1.77)	0.997	not a control; FAIL	1.40 (1.08-1.84)
All disabilities (claim-relevant; 1,448)	1.09 (1.00-1.19)	0.980	not a control; PASS	1.14 (1.00-1.27)

Notes. U.S.-alone posteriors of the core joint likelihood on each outcome's observed cells (suppressed, unpublished and quality-flagged cells excluded, none imputed). Width is the interval's log width relative to autism's. The future+3 placebo shifts the PCV path three cohorts later on the same cells; FAIL means some kernel setting excludes one. All disabilities is read at power 20 and contains autism. Source: `3326_us_core_negative_control_outcomes.md` and `.csv`.

Only hearing impairment is precise enough to count, and it leans like autism, so the stated verdict is MIXED: the rule establishes neither that the autism lean is shared state-cohort drift nor that it is specific to autism. The complete-series sensitivity also reads MIXED. The controls' position relative to autism is not stable across cell rules: on complete series traumatic brain injury is 0.83 (0.42-1.59) on 222 cells, below autism. The DDD fits (autism relative to each control) restate the comparison on the control's cells and are not a second line of evidence. The future+3 placebo fails for intellectual disability, speech or language, specific learning disability and other health impairment, so their PCV leans are not interpretable as vaccine effects; autism and all four controls pass it. The controls also carry nominal

rotavirus-2006 leans where autism gives 0.95. Suppression is informative, so rare categories are observed mainly in large states and higher-count years, and the controls are less precise than autism by construction.

S63.3 U.S. injury controls

The U.S. injury task swaps injury outcomes into six autism designs. Of ten primary estimates, one is a BIAS SIGNAL, six are CONSISTENT, two are UNINFORMATIVE and one is NOT VALIDATED because the autism refit did not reproduce the published SLAITS estimate.

Table S63-B. U.S. injury negative controls, primary estimates.

Design (outcome)	Cells or children	Injury r (95% interval)	Autism on the same cells	Class
Core joint likelihood, observed cells (IDEA traumatic brain injury)	472	1.65 (0.77-4.29)	1.11 (0.80-1.49)	UNINFORMATIVE
Core extended to single ages 6-11 (IDEA traumatic brain injury)	977	0.84 (0.46-1.35)	1.06 (0.91-1.22)	UNINFORMATIVE
NSCH legacy, PCV 1+ dose (brain injury or concussion, ever)	120,005	1.12 (1.02-1.23)	1.10 (0.93-1.32)	BIAS SIGNAL, harm direction
NSCH modern, combined seven-series	77,105	0.85 (0.68-1.06)	1.00 (0.84-1.20)	CONSISTENT
NSCH modern, PCV 4+ doses	77,105	0.81 (0.63-1.04)	0.92 (0.75-1.13)	CONSISTENT
NSCH modern, rotavirus	77,105	0.90 (0.72-1.12)	0.93 (0.79-1.10)	CONSISTENT
SLAITS 2007 and 2011, PCV7 rollout (brain injury or concussion)	79,965	1.01 (0.74-1.38)	0.96 (0.07-14.14)	NOT VALIDATED
SLAITS 2003 and 2007, ages 0-5 (injury needing medical care)	47,169	1.01 (0.94-1.08)	1.11 (0.55-2.26)	CONSISTENT
California CASEMIS PCV7 jump at age 11 (traumatic brain injury)	19	1.05 (0.32-3.52)	1.03 (0.57-1.85)	CONSISTENT
California CASEMIS rotavirus 2006 jump at age 8 (traumatic brain injury)	16	1.13 (0.34-3.79)	1.22 (0.58-2.55)	CONSISTENT

Notes. Core rows are U.S.-alone posteriors with 95 percent credible intervals; survey rows use state-clustered Wald intervals and CASEMIS rows placebo-calibrated intervals. The modern NSCH combined seven-series and PCV 4+ rows fail their future+3 placebo. Source: `3331_injury_negative_controls_us.md` and `.csv`.

The bias signal comes from a separate outcome system in the same children: in the NSCH dose-response design, parent-reported brain injury or concussion rises 1.12 (1.02-1.23) per ten points of PCV coverage, beside autism 1.10 (0.93-1.32), and its future+3 placebo passes. Among secondary estimates, which never enter a reading, traumatic brain injury in the core likelihood gives 2.37 (1.09-4.58) for rotavirus 2006. Numerically unstable SLAITS standard errors motivated a post-hoc full-rank refit that is reported but not used for classification. The published SLAITS estimate that this frozen audit compares against has since been replaced by a full-rank refit with a corrected birth-cohort rule for the latest survey wave (Section S14); the audit was not re-run. Survey items are parent reports whose wording changed in 2020, and suppression of small IDEA cells selects on the outcome.

Round two fitted child deaths from unintentional injuries, taken from the NCHS multiple-cause microdata for 2000-2004, to the core-v2 U.S. likelihood on state-by-birth-cohort Lexis cells at ages 0-5, with the exposure paths, kernels, priors and seeds unchanged. No autism outcome sits on these cells, so the width reference is the core-v2 autism posterior of the same product. The primary PCV estimate, on 1,107 cells with 12,715 deaths after 88 zero-count cells were dropped, is 2.59 (1.15-5.76) with $P(r > 1) = 0.989$. It is UNINFORMATIVE because its interval is 4.44 times as wide as autism's 1.17 (0.97-1.39) on the log scale; that it excludes one is reported as a flag and does not change the class. Its future+3 placebo fails in the opposite direction, 0.18 (0.10-0.39), so on these cells the machinery returns associations that are not specific to PCV timing. RotaShield gives 0.82 (0.17-3.13), CONSISTENT but prior-dominated, against autism 0.87 (0.36-1.83). Rotavirus 2006 is not identified, because the core assigns no rotavirus-2006 exposure to the 1997-2004 cohorts that make up every death cell. Among secondary fits, PCV gives 2.69 (1.23-5.76) at ages 1-5 and 1.49 (0.66-3.60) for all external causes, both with failing placebos, and 1.08 (0.50-2.78) in age bands 0, 1-2 and 3-4, whose placebo passes; the head-injury fit is degenerate (pooled effective sample size 1.0) and carries no information. On the core's own age-8 state-cohort units (351 of 353), autism relative to injury deaths of the same state-cohorts gives PCV 1.03 (0.69-1.52) and RotaShield 1.31 (0.48-2.84), a descriptive DDD. Deaths at ages 0-5 are flows that share no age, year or count date with the autism counts, so they can detect state-cohort biases that act through population composition,

poverty or access to care, not biases that act through identification. Public death files from 2005 on carry no state, so the rotavirus-2006 cohorts cannot be tested this way.

S63.4 Brazil

The Brazilian Census design is the pooled maternal DTaP context: 108 state-by-age-band cells, with SIH/SUS injury admissions for 2021-2023 per Census 2022 population replacing Census-reported autism. Autism on these cells is 0.77 (0.51-1.14).

The all-injury composite gives 0.51 (0.29-0.83), a bias signal in the protective direction, and falls give 0.38 (0.15-0.86), also a bias signal. Head injury 0.47 (0.18-1.10), fractures 0.73 (0.40-1.29), burns 1.47 (0.65-2.99), poisoning 0.67 (0.24-1.65) and road traffic 1.13 (0.56-2.08) are CONSISTENT; foreign body, dog bite and drowning have zero-count cells and are UNINFORMATIVE. Six of nine components are informative. The stated reading is DESIGN BIAS DETECTED ($r < 1$), and it holds with age at admission and with 2022 alone. The future+3 placebo fails for all injuries, 0.63 (0.43-0.90), and for head injury, burns and falls, while passing for autism, 0.84 (0.64-1.14). The harm-injection gate passes for every outcome (noise-free median at an injected ratio of 2: autism 2.09, all injuries 1.92), so the design responds to real effects; the DDD for autism relative to all injuries is 1.52 (0.96-2.38).

The same Census cells carry four report-only product designs, which are not independent tests because they share cells and counts. Hepatitis A gives autism 2.17 (1.36-3.98), whose own future+3 placebo fails at 1.85 (1.23-3.09), and all injuries 2.01 (1.15-4.23): DESIGN BIAS DETECTED ($r > 1$). PCV10 with MenC gives all injuries 0.44 (0.28-0.75) against autism 1.19 (0.76-2.05): DESIGN BIAS DETECTED ($r < 1$). Rotavirus reads NO BIAS DETECTED (all injuries 0.61, 0.35-1.47) and MMRV reads MIXED (1.25, 0.80-2.66, with a harm-direction head-injury signal). Admissions are flows while Census autism is a stock, denominators are population rather than person-time, admissions are SUS-only authorizations rather than persons, and 2021 falls in the pandemic period.

Round two replaced admissions with SIM deaths from unintentional injuries (V01-X59; death years 2021-2023, age at death) on the same 108 cells, exposure path, kernels and prior, under a rule written before any death count was seen. The composite gives 0.36 (0.15-0.79), $P(r > 1) = 0.005$, a bias signal in the protective direction, against autism 0.77 (0.51-1.14); its interval is 2.08 times autism's on the log scale. Drowning, 0.75 (0.27-1.90), is the only informative component and is CONSISTENT; road traffic, falls, fire and accidental poisoning have zero-death cells and are UNINFORMATIVE. The reading is again DESIGN BIAS DETECTED ($r < 1$), and it holds for death years 2020-2024 (0.36, 0.15-0.77). The composite's future+3 placebo fails, 0.31 (0.15-0.57), while autism's passes, and autism relative to injury deaths is 1.92 (0.97-3.57). The harm-injection gate passes (noise-free median 1.38 at an injected ratio of 2), but its no-effect median is biased to 0.69 and the plasmode check fails (0.63). Deaths come from a different register from admissions and depend less on SUS hospital access; the two bias signals are reported side by side, not combined. Among the report-only products on the same cells, unintentional injury deaths give bias signals for rotavirus, 0.37 (0.19-0.86), and PCV10 with MenC, 0.26 (0.15-0.54), both read DESIGN BIAS DETECTED ($r < 1$), the rotavirus reading changing from round one's NO BIAS DETECTED; MMRV reads NO BIAS DETECTED (1.22, 0.61-1.96) and hepatitis A MIXED (0.75, 0.38-1.74). Ages are ages at death rather than the Census cohort ages, and 2021 falls in the pandemic period.

S63.5 United Kingdom and northern Europe

The UK task fits WHO Mortality Database injury deaths for 2006-2024 on cell sets closest to the English, Scottish and Northern Irish rotavirus autism cells (S) and on early-age extensions (E). The injury cells are never the autism cells, so the width reference is the published autism posterior. Its rule adds thinness (fewer than 100 deaths or more than 20 percent zero cells) and prior dominance to the width criterion. Of six primary estimates (all unintentional injury deaths), five are UNINFORMATIVE and one is CONSISTENT, Scotland E at 0.87 (0.32-2.31), whose class is not stable under a wider prior; none is a bias signal, and 73 of 75 secondary estimates are UNINFORMATIVE. England S gives 0.57 (0.25-1.22) with a failing future+3 placebo and harm gate. On a 23-cell English bridge where autism and injury share one layout, autism is 0.92 (0.48-1.56) and injury deaths 0.48 (0.18-1.15), a secondary CONSISTENT reading.

Round two acquired injury series that sit on or beside the UK autism cells: HES admissions by primary diagnosis and by external cause and STATS19 road casualties for England, PHS admissions and council-level STATS19 casualties for Scotland, PSNI casualties for Northern Ireland, and injury admissions at ages 0-4 by upper-tier local authority for the report-only English MMR design. Each control is fitted on a set S paired with autism refitted on the same cells and on an extension E whose reference is the published autism posterior; every autism block reproduced its published posterior

first. Of 14 primary estimates, 6 are CONSISTENT, 7 UNINFORMATIVE and none a bias signal; the English MMR extension collapses to its S set (Table S63-C). Round one had no informative control for England. STATS19 casualties on the English rotavirus block restricted to census years 2011-2024 (112 cells, 100,332 casualties) give 1.00 (0.87-1.16) against autism 0.97 (0.89-1.06) on the same cells, an interval 1.62 times as wide on the log scale, and the extension gives 1.06 (0.94-1.19). HES admissions give 0.99 (0.57-1.57) by primary diagnosis and 0.98 (0.53-1.60) by external cause on 23 and 21 band-by-year cells, against autism 0.89 (0.53-1.38) and 0.91 (0.52-1.43) on the same cells; their extensions are UNINFORMATIVE against the narrow published interval. One of 53 secondary rows is a bias signal: HES poisoning admissions on set S, 0.41 (0.19-0.85), in the protective direction. Northern Irish casualties give 1.02 (0.37-2.04) against autism 0.65 (0.29-1.32) on the same 22 cells. Scotland remains untestable on the school-age cells: PHS admissions, 0.92 (0.26-3.05), and STATS19 casualties on the identical 490 autism cells, 0.75 (0.22-2.42), are prior-dominated, although the PHS extension over ages 0-14 is CONSISTENT at 1.11 (0.42-2.75). The English MMR control, 1.29 (0.38-4.49), is UNINFORMATIVE. Every fitted primary future+3 placebo passes, and the harm-injection gate passes on every English and Northern Irish row (plasmode medians 1.86-2.13 at an injected doubling). HES counts are finished consultant episodes, not persons, and are affected by the move of same-day emergency care to the Emergency Care Data Set in 2021/22-2023/24; STATS19 records casualties where the collision occurred, and 2020 is a pandemic year.

Table S63-C. UK round-two injury controls, primary estimates.

Design, control (set; cells)	Injury r (95% CrI)	Autism reference	Width ratio	Class
England rotavirus, HES diagnosis (S; 23)	0.99 (0.57-1.57)	0.89 (0.53-1.38)	1.07	CONSISTENT
England rotavirus, HES diagnosis (E; 38)	0.96 (0.69-1.29)	published 0.94 (0.86-1.02)	3.78	UNINFORMATIVE
England rotavirus, HES external cause (S; 21)	0.98 (0.53-1.60)	0.91 (0.52-1.43)	1.09	CONSISTENT
England rotavirus, HES external cause (E; 35)	0.98 (0.69-1.34)	published 0.94 (0.86-1.02)	3.99	UNINFORMATIVE
England rotavirus, STATS19 casualties (S; 112)	1.00 (0.87-1.16)	0.97 (0.89-1.06)	1.62	CONSISTENT
England rotavirus, STATS19 casualties (E; 195)	1.06 (0.94-1.19)	published 0.94 (0.86-1.02)	1.38	CONSISTENT
Scotland rotavirus, PHS admissions (S; 306)	0.92 (0.26-3.05)	1.23 (0.31-4.45)	0.93	UNINFORMATIVE, prior-dominated
Scotland rotavirus, PHS admissions (E; 960)	1.11 (0.42-2.75)	published 1.55 (0.38-5.31)	0.70	CONSISTENT
Scotland rotavirus, STATS19 casualties (S; 490)	0.75 (0.22-2.42)	1.55 (0.38-5.33)	0.91	UNINFORMATIVE, prior-dominated
Scotland rotavirus, STATS19 casualties (E; 512)	0.75 (0.22-2.42)	published 1.55 (0.38-5.31)	0.91	UNINFORMATIVE, prior-dominated
Northern Ireland rotavirus, PSNI casualties (S; 22)	1.02 (0.37-2.04)	0.65 (0.29-1.32)	1.11	CONSISTENT
Northern Ireland rotavirus, PSNI casualties (E; 24)	1.16 (0.48-2.17)	published 1.01 (0.88-1.16)	5.47	UNINFORMATIVE
England MMR (report-only), local-authority admissions at ages 0-4 (S; 1,020)	1.29 (0.38-4.49)	published 0.53 (0.36-0.96), other cells	2.51	UNINFORMATIVE
England MMR (report-only), local-authority admissions at ages 0-4 (E)	not fitted	none	none	collapses to S

Notes. Width ratio is the injury interval's log width relative to the reference's: autism refitted on the same cells for S rows (for the Scottish STATS19 S row, the published fit on the identical cells; for the English MMR row, which has no same cells, the published fit) and the published autism posterior for E rows. A row is UNINFORMATIVE if the ratio exceeds 2.5, if it is thin or if it is prior-dominated (log width at least 0.8 x 2.7171). Posterior medians and 95 percent credible intervals from grid quadrature. Source: `3380_injury_negative_controls_uk_round2.md` and `.csv`.

In round one no European pooled context (Sweden PCV, Germany PCV, Czechia MMR) had injury data on its own cells; round two supplies them (Section S63.12). The round-one European task therefore pairs injury outcomes with report-only blocks whose paired outcomes are not autism: Norwegian primary-care injuries against symptom codes P70-P99, and Estonian injury diagnoses S00-T98 against F00-F99 care for hepatitis B. Norway gives 1.25 (0.36-2.21) against a paired 0.50 (0.18-1.63) and Estonia 1.87 (0.64-3.50) against 1.08 (0.15-2.99); both read CONSISTENT, and both lean above one. An injected doubling moves the Norwegian injury estimate from 1.25 only to 1.69, so a

CONSISTENT reading here is weak evidence that bias is absent. A post-hoc analysis finds the Estonian injury lean identified mainly off Tallinn.

S63.6 Chile, Colombia and Tanzania

Chile's PCV10 discharge design (40 national band-by-year cells; a report-only autism block that fails the plasmode gate) was run under two stated rules. All injuries give 0.95 (0.54-1.54), head injury 1.07 (0.62-1.76) and fractures 1.00 (0.60-1.66), all CONSISTENT, against autism 0.50 (0.17-1.57). The readings are NO BIAS DETECTED under the first rule and CONSISTENT under the second, which counts the system once. Consistency here is weak evidence. An injected doubling on all injuries returns 1.59 (0.81-2.48) on the observed data, still CONSISTENT, and an autism-sized ratio of 0.5 returns 0.61 (0.34-1.12), also CONSISTENT. The autism median lies below every primary control's lower limit (0.54, 0.62 and 0.60), and the design's noise-free null median for autism is 0.555 against the observed 0.499, so the autism estimate sits near the value the design returns when nothing happens. The width criterion cannot fire here, because 2.5 times the autism interval's log width exceeds the prior's own width, and the thin-count threshold was set with counts in view.

Colombia's varicella school-census design (the pooled context; 497 to 504 department-by-level-by-year cells) gives hearing disability 0.98 (0.66-1.49), motor or physical disability 0.87 (0.67-1.28) and visual disability 0.90 (0.58-1.27), all CONSISTENT, against autism of 0.75 (0.49-1.04) and 0.77 (0.51-1.08) on the same cells; the system reading is CONSISTENT. These categories are specificity outcomes rather than injuries. In round one, injury controls could not be tested for this context: the census has no injury category and Colombian mortality rows lack subnational codes. Tanzania's schedule design (49 grade-by-year cells) uses albinism, a genetic condition, because no injury outcome exists; it gives 1.00 (0.69-1.28) against autism 1.12 (0.54-1.69), CONSISTENT, with a DDD of 1.06 (0.59-1.61). The autism median lies inside the albinism interval, so this control speaks only against artifacts outside 0.69-1.28. An older regional first-difference design had shown nominal positive albinism associations (for example DTP1 1.031, 1.013-1.049, per ten points), which is why the albinism control is not treated as proof of a clean design.

Round two placed INMLCF medico-legal records of non-fatal transport injuries, by department of the event, life-cycle band and year, on the Colombian block's own region-by-level-by-year cells (500 of 504 cells; 13,745 records), with DANE band populations as denominators. The primary control gives 1.36 (0.99-1.93), $P(r > 1) = 0.973$, CONSISTENT with a lower limit of 0.991, against autism 0.76 (0.51-1.07) on the same cells; autism relative to transport injuries is 0.63 (0.39-1.02). The class is borderline. Counting all cells with a 0.5 continuity correction gives 1.39 (1.03-2.02), and transport plus accidental injuries give 1.33 (1.01-1.96), both harm-direction bias signals, and the primary is a bias signal in 5 of the 12 kernel-by-prior settings, including all three linear-only settings. Dropping the 2020-2021 school years moves the control to 1.18 (0.78-1.80) and autism on the same 335 cells from 0.76 (0.51-1.07) to 1.31 (0.89-1.98), so the Colombian autism estimate itself depends on the lockdown years. Section S62.8 carries a lockdown-free Colombian curve into the synthesis (file 3371): on 336 cells, keeping the one cell that the transport rule drops, autism gives 1.32 (0.90-1.99), and the shared mean moves from 0.957 to 1.045. Every Colombian future+3 placebo passes. The descriptive notes on these sensitivities were written after the results were read, and none changes the class. INMLCF records only injuries that reach a forensic examination, by the department where the event happened rather than the department of the pupil's school, and its life-cycle bands only approximate the school levels.

S63.7 Japan

Japan's special-support class survey supplies sensory and physical class placements rather than injuries. The headline design is the Japanese encephalitis block that represents this system in the hierarchy. On 803 prefecture-by-grade-by-year cells the pooled hearing, low-vision and physical classes give 0.91 (0.62-1.31) against autism 1.02 (0.87-1.18), CONSISTENT, with an interval 2.48 times autism's against the limit of 2.5. That interval contains the whole autism interval, so the control cannot rule out a non-specific bias of that size, and its class is CONSISTENT in only 6 of 12 kernel-by-prior combinations (a descriptive flag that does not change the class). Hearing classes alone give 0.55 (0.32-0.98), UNINFORMATIVE because thin, with a failing future+3 placebo that points to a Hokkaido-specific trend. The synthetic-control version is UNINFORMATIVE, and four national MEXT jump reduced forms read CONSISTENT; in those reduced forms the controls move with autism and fail the shifted-path placebo as often as autism does (26 of 40 rows each).

The Japan MMR design, whose exposure is the Kohoku ward coverage path, gives 0.86 (0.46-1.61) for national unintentional injury deaths across its nine birth cohorts, CONSISTENT, against autism 0.85 (0.28-2.32) on its own

Kohoku cells. The injury estimate sits where autism's does, so the control neither supports nor undercuts reading the Kohoku estimate as an MMR effect. Round two tried Yokohama city deaths, the city that contains Kohoku ward, but its cause-specific tables begin in 2000, so the only usable cells are cohort-cumulative deaths at ages 15-19 of the nine cohorts, with national births as denominators. Accident deaths give 1.00 (0.41-2.21), UNINFORMATIVE because the cohort cells are thin (minimum 6.2 deaths, against the 16 the rule requires).

S63.8 Hong Kong

The Hong Kong PCV design (the pooled context; 60 grade-by-school-year cells) gives autism 0.89 (0.66-1.16). With autism replaced by other special-education categories, physical disability gives 1.78 (1.23-2.90) and physical plus visual impairment 1.50 (1.09-2.22), both bias signals in the harm direction; the composite of physical, visual and hearing impairment, 1.15 (0.90-1.47), and hearing impairment, 0.99 (0.72-1.32), are CONSISTENT, and visual impairment, 2.01 (0.56-6.06), is UNINFORMATIVE. One informative bias signal fires the system reading, BIAS SIGNAL. This rule was written with the deciding estimates already known, so the Hong Kong reading is not a blind test. The physical-disability interval excludes one at a Bonferroni 99 percent level across the five controls (1.06-3.47); the physical-plus-visual interval does not (0.97-2.57). Physical disability passes its future+3 placebo (0.69, 0.38-4.21) but its two-sided placebo family fails in 3 of 20 member-by-setting intervals; physical plus visual fails its future+3 placebo (0.58, 0.41-0.86) and 5 of 20 family intervals.

A null calibration simulates the design with no effect. With cell count noise only, the 95 percent interval excludes one in 0.0-4.0 percent of simulations, near nominal; adding birth-cohort deviations raises this to 13.5 percent for autism, 8.8 percent for physical plus visual and 10.5 percent for physical disability. The design's intervals therefore under-cover for autism as well as for the controls, and under the cohort-deviation null the observed physical-disability estimate lies at the 12.0 percent two-sided tail. On the log scale the physical-disability lean is 0.58 and the physical-plus-visual lean 0.41, both smaller than the context bias SD of 0.70 that the hierarchy already assigns to Hong Kong PCV. Autism relative to the controls runs from 0.55 to 0.98, so no single bias correction follows. The older frequentist screen of the same data reported physical disability at 1.049 (1.016-1.084) per ten points of coverage (Section S48.1).

Hong Kong injury deaths (158 deaths at ages 1-9 in 2002-2018, on cells that differ from the autism cells) give 0.89 (0.26-3.74), an interval 4.67 times autism's and dominated by the prior; the reading is UNINFORMATIVE, so injury deaths cannot test this design.

Round two supplies the first informative injury control for Hong Kong, under a rule written before any fit: police-reported road-traffic casualties at ages 5-19 in 2009-2024 (48 band-by-year cells, 22,323 casualties), with the autism model and the exposure path mapped from birth cohort. They give 0.66 (0.48-0.92), $P(r > 1) = 0.008$, a bias signal in the protective direction, with an interval 1.15 times the width of autism's on its own cells. The class is stable under every kernel setting and prior, and the future+3 placebo passes (0.88, 0.56-1.64). Excluding 2020-2022 gives 0.74 (0.57-0.94), still a bias signal; the 2013-2021 window gives 0.98 (0.59-1.63) and pedestrian casualties in 2014-2024 give 0.76 (0.52-1.15), both CONSISTENT. The signal lies on the same side of one as autism's 0.89 (0.66-1.16), and autism relative to road casualties is 1.34 (0.86-2.05), a contrast of independent posteriors because the casualty cells (age bands by calendar year) are not the autism cells (grades by school year). Casualties are recorded by accident location, not residence. The Hong Kong design therefore returns non-specific leans in both directions: harm-direction leans in two special-education categories and a protective-direction lean in road casualties.

S63.9 Korea and Taiwan

Korea's special-education design (918 province-by-grade-by-year cells; a report-only block that fails four pooling gates) gives autism 0.53 (0.26-0.80). The sensory-and-physical composite gives 0.90 (0.56-1.50), visual impairment 1.13 (0.41-3.17), hearing impairment 0.98 (0.56-2.10) and physical impairment 0.89 (0.53-1.78), all CONSISTENT; the reading is CONSISTENT. The design does not reproduce autism's protective lean in these categories, and autism's own future+3 placebo fails (1.57, 1.08-3.98) while none of the four controls' placebos does. Korean injury deaths are national series in five-year age bands. On the cells matched to the autism block every outcome is prior-dominated and the reading is UNINFORMATIVE; an injected doubling on these cells is recovered as 0.95. A secondary design over ages 1-19 in 2008-2023 reads CONSISTENT (external causes 0.98, 0.44-2.69; unintentional injuries 0.93, 0.40-2.64), but it places only 0.065 and 0.092 of its posterior mass below the autism estimate of 0.53, so it could not have excluded a bias as large as that estimate.

Taiwan's varicella design on its special-education cube (720 county-by-grade-by-year cells; report-only because its autism future+3 placebo fails) gives the sensory-and-physical composite 0.89 (0.73-1.04) against autism 1.07 (0.87-1.44), CONSISTENT; visual 1.08 (0.77-1.80), hearing 0.88 (0.68-1.09) and physical 0.88 (0.68-1.09) are also CONSISTENT, and the verdict is NO BIAS DETECTED. The future+3 placebo fails for the composite (0.85, 0.68-0.99) and for physical impairment (0.77, 0.57-0.94), as it does for autism (1.40, 1.12-2.20) in the opposite direction, so the null actual-cutoff controls do not certify the design. Pretrend and harm-injection gates were added after the rule and do not change the verdict. Round two fitted MOHW accident deaths by county of residence to the same cube (628 of 720 cells with deaths; 5,676 deaths): 0.91 (0.48-2.05) against autism 1.07 (0.86-1.46) on the same cells, UNINFORMATIVE because the interval is 2.74 times as wide on the log scale. All six Taiwanese death fits are UNINFORMATIVE (2.54-3.06 times autism's width) and none excludes one; their future+3 placebo passes in all six fits for the control and fails in all six for autism on the same cells.

S63.10 Injury deaths across national designs and German discharges

The mortality task refits eight national autism designs on WHO Mortality Database injury deaths. Only Chile's cells are identical to its autism cells; elsewhere the injury cells are national series that differ from the autism cells, so the autism-relative contrast is not a DDD. Across the eight contexts the readings are 5 CONSISTENT, 0 BIAS SIGNAL and 3 UNINFORMATIVE: Sweden PCV 0.71 (0.32-1.62), Germany PCV 1.01 (0.50-1.81), Japan MMR all injuries 0.89 (0.48-1.67), Brazil maternal DTaP 1.08 (0.79-1.40) on differently aligned cells, and Chile PCV 0.79 (0.40-1.49) are CONSISTENT; England rotavirus 0.57 (0.25-1.22) and Scotland rotavirus 0.91 (0.25-3.25) are UNINFORMATIVE; Colombia varicella is not fitted because no age-specific population is held. The German reading uses grouped ages 1-4 in 2000-2009, under an implementation note written after the rule but before the affected fits, because WHO publishes the German population at ages 1-4 only as a group in those years; under the literal rule the specified single-age cells are not fittable, and the readings would be 4 CONSISTENT, 0 BIAS SIGNAL and 4 UNINFORMATIVE. The English matched cells cannot test the English autism estimate: their noise-free no-effect median is 0.55, against 0.99 on the autism cells. Benchmarks without an autism model give U.S. PCV7 in 2000 0.99 (0.66-1.34), English MenB in 2015 1.21 (0.78-1.84), Japanese hepatitis B in 2016 1.04 (0.73-1.48) and Japanese rotavirus in 2020 0.87 (0.48-1.53), all CONSISTENT, while the English hexavalent hepatitis B step in 2017 gives unintentional injury deaths 2.21 (1.17-3.94), a bias signal on cells that include the pandemic years (Section S60.5).

In round one German injury discharges were held only for 2024 (round two added 2000-2023; Section S63.12), a cross-section of post-programme cohorts that tests the Land-by-age-band arm of the PCV design, not the 2006 introduction. All seven primary controls are UNINFORMATIVE under the stated prior-dominance criterion (their literal classes would be CONSISTENT), so the reading is UNINFORMATIVE in both the merged-region and 16-Land designs. The cross-section cannot return a true effect: an injected doubling returns 0.72 and 0.64 in the two designs. A secondary single-code control, shoulder and upper-arm fractures, gives 0.22 (0.08-0.65) in the 16-Land design, a protective-direction flag consistent with the design's downward pull (its noise-free null median is 0.21).

S63.11 Canada

Round one held no control for the Canadian rotavirus contexts. Round two uses asthma from the Canadian Chronic Disease Surveillance System (CCDSS), which is published at ages 1-19 for every province-year of the design excluding Quebec, so all 166 of its cells are kept. No established pathway from rotavirus vaccination to asthma is known; the download request chose asthma as the least-bad CCDSS condition because no public Canadian injury series fits the cells, not because its implausibility had been argued. Asthma incidence gives 0.71 (0.48-1.06), P(r > 1) = 0.036, CONSISTENT, against the reproduced autism estimate of 0.99 (0.32-1.89); its interval is 0.45 times autism's on the log scale, and its future+3 placebo, pretrend check and harm-injection gate pass. The class is CONSISTENT in 10 of 12 kernel-by-prior settings; the two using the rho 0.85 kernel give bias signals, 0.65 (0.44-0.96) and 0.64 (0.39-0.94). Asthma prevalence, a sensitivity, gives 1.08 (0.86-1.39) with a failing future+3 placebo (1.26, 1.03-1.60), and autism relative to asthma incidence is 1.34 (0.47-2.20). Quebec's autism cells are INSPQ age-band prevalence, while the CCDSS publishes Quebec only for ages 1-19, so no identical-cell control exists; a report-only analogue on 24 Quebec CCDSS cells gives asthma 0.92 (0.25-3.21) against autism 1.37 (0.33-5.34) on the same cells. CIHI injury tables do not cross-classify province by age and were not fitted. Asthma shares the CCDSS case-finding of physician billing and hospital codes with the Canadian autism outcome, so it can detect case-finding drift aligned with the launch dates but not autism-specific confounding such as an expansion of autism services; an indirect pathway, such as fewer gastroenteritis visits at which asthma could first be coded, would be small and in the r < 1 direction. Asthma is not an injury and cannot enter the injury calibration set of Section S63.13.

S63.12 Sweden, Germany and Czechia

In round one the Swedish and German PCV designs could be tested only against national WHO death series on other cells and, for Germany, a 2024 discharge cross-section (Section S63.10), and the Czech MMR design had no control. Round two (file 3381) placed care-register injuries on or beside their own cells: Socialstyrelsen patient-register counts, unique patients per year and code in inpatient and specialised outpatient care, on the 756 county-by-age-band-by-year cells of the Swedish PCV block; Destatis hospital discharge episodes on the 36 cells of the frozen German PCV block and, beside the reading, on the 825 cells of the report-only rebuilt German panel; and NRHZS counts of unique patients in publicly reimbursed care for the Czech MMR model. The rule is the round-one German rule of file 3348: a control is UNINFORMATIVE if it is not fittable, is thinner than the autism series it checks, or is more than 2.5 times as wide on the log scale as autism refitted on the same cells; otherwise it is a BIAS SIGNAL if its interval excludes one; otherwise UNINFORMATIVE if prior-dominated (log width at least 0.8 x 2.7171); and otherwise CONSISTENT. Each context is read from seven primary controls: all injuries (S00-T98 excluding T78 and T80-T88), head injuries, intracranial injury, fractures, burns, poisoning and foreign body (Table S63-E). Every autism block first reproduced its published posterior exactly.

Table S63-E. Round-two European injury controls on the pooled designs, primary estimates.

Control	Sweden PCV (756 cells)	Germany PCV, frozen (36 cells)	Czechia MMR, closest cells (462)	Germany PCV rebuilt, report-only (825 cells)
Autism (width reference)	1.01 (0.79-1.56)	1.03 (0.27-3.96)	0.91 (0.27-3.37)	0.49 (0.21-1.08)
All injuries	1.06 (0.98-1.13)	1.09 (0.31-3.91)	1.19 (0.32-4.48)	1.12 (0.94-1.32)
Head injuries	1.04 (0.94-1.15)	0.98 (0.28-3.51)	1.08 (0.30-4.10)	1.09 (0.87-1.36)
Intracranial injury	1.01 (0.91-1.12), thin	1.02 (0.28-3.74)	1.05 (0.29-3.87)	1.24 (0.94-1.62)
Fractures	1.04 (0.97-1.11)	1.22 (0.33-4.37)	1.12 (0.31-4.18)	1.10 (0.91-1.32)
Burns	1.16 (0.92-1.44), thin	0.84 (0.23-3.20)	1.10 (0.31-3.97)	0.69 (0.40-1.18)
Poisoning	1.38 (0.93-2.01), thin	0.68 (0.18-2.89)	1.03 (0.29-3.82)	1.28 (0.77-2.13)
Foreign body	1.18 (0.96-1.34), thin	0.90 (0.23-3.70)	1.02 (0.28-3.81)	1.56 (0.89-2.43)
Reading	CONSISTENT	UNINFORMATIVE	UNINFORMATIVE	CONSISTENT, beside the reading

Notes. Posterior medians with 95 percent credible intervals. Autism is refitted on each design's cells, except in Czechia, where the width reference is autism on its own 847 cells of 2019-2024, because the injury cells (ages 5-9) are the closest available rather than the autism cells (ages 4-6 and 7-10). Rows marked thin are UNINFORMATIVE because they are sparser than the autism series. Every control on the frozen German cells and on the Czech cells, like the frozen German and Czech autism posteriors, is prior-dominated and therefore UNINFORMATIVE, although its literal class

would be CONSISTENT. Swedish burns and poisoning are fitted on 612 and 684 cells after the balanced-series rule dropped 8 and 4 county-by-band series. Source: `3381_injury_negative_controls_europe_round2.md` and `.csv`.

Sweden now has an informative injury control on a European block's own cells. All injuries give 1.06 (0.98-1.13), P(r > 1) = 0.932, against autism 1.01 (0.79-1.56) on the same cells; the interval is 0.21 times autism's on the log scale, and head injuries, 1.04 (0.94-1.15), and fractures, 1.04 (0.97-1.11), are also CONSISTENT. Intracranial injury, burns, poisoning and foreign body are UNINFORMATIVE because they are thinner than Sweden's dense autism series, so the reading, CONSISTENT, rests on three controls. The all-injury class is CONSISTENT under every alternative kernel and prior, its future+3 placebo passes (1.00, 0.89-1.07), the harm-injection and plasmode checks pass for every primary control, and autism relative to all injuries is 0.97 (0.74-1.56). Two thin secondary fracture codes have intervals that exclude one, lumbar or pelvic fracture 2.02 (1.09-3.58) and shoulder or upper-arm fracture 1.12 (1.02-1.21); both are UNINFORMATIVE under the rule, and the shoulder code fails its future+3 placebo. External-cause series on three-year age bands, reported beside the reading, agree: all accidents give 1.03 (0.98-1.08) in inpatient and outpatient care and 0.99 (0.89-1.06) in inpatient care alone, whose future+3 placebo fails.

The frozen German block cannot be tested with useful precision. On its 36 cells every primary control is prior-dominated, as autism is at 1.03 (0.27-3.96): all injuries give 1.09 (0.31-3.91), the seven controls' log widths are 0.94-1.04 times autism's, and the reading is UNINFORMATIVE. No future+3 placebo is testable on these cells. On the report-only rebuilt panel (825 merged-region-by-year-by-age-band cells), where autism is 0.49 (0.21-1.08), all seven primary controls are CONSISTENT (all injuries 1.12, 0.94-1.32) and all 21 fitted future+3 placebos pass; among its secondary codes, other external causes including drowning give a protective-direction bias signal, 0.63 (0.37-0.99). That panel's autism block fails its plasmode check (an injected doubling is read back as 0.70) and remains a report-only sensitivity, reported beside the reading and never in place of it.

Czechia has no identical cells. NRHZS publishes five-year age bands only, so the controls sit on the closest cells, 77 districts at ages 5-9 in 2019-2024 (462 cells), against autism cells at ages 4-6 in 2016-2024 and 7-10 in 2020-2024, with the exposure re-averaged over each band's five cohorts. The width reference, autism on its own 2019-2024 cells, is itself prior-dominated at 0.91 (0.27-3.37). Every primary control is prior-dominated (all injuries 1.19, 0.32-4.48), the reading is UNINFORMATIVE, and no DDD can be formed. The all-injury future+3 placebo failed to fit in the production routine and is NOT TESTABLE (a post hoc refit gives 1.18, 0.31-4.50); the 16 fitted future+3 placebos pass.

These are care-use counts, not incidence: the German counts are discharge episodes rather than persons, Swedish composites sum per-code patient counts, so a child with two codes counts twice, and the Czech counts are patients with the code in the main position. The German counts for 2000-2023 come from pruned GENESIS table states whose parent rows were verified to carry true totals on held 2024 cells, and external-cause coding in Swedish outpatient care varies by county and year.

S63.13 Readings by system

Table S63-D. Stated negative-control readings by system.

Task and design (autism context status)	Primary control: r (95% interval)	Autism on the same or own cells	Stated reading
U.S. IDEA categories, core PCV (pooled)	hearing impairment 1.41 (1.05-1.91)	1.17 (0.97-1.39)	MIXED
U.S. injuries, six designs (pooled core and report-only)	NSCH brain injury 1.12 (1.02-1.23)	1.10 (0.93-1.32)	1 BIAS SIGNAL, 6 CONSISTENT, 2 UNINFORMATIVE, 1 NOT VALIDATED
U.S. injury deaths, NCHS microdata, core-v2 paths (pooled core; different cells; round two)	PCV: unintentional injury deaths 2.59 (1.15-5.76)	1.17 (0.97-1.39)	UNINFORMATIVE (interval excludes 1; placebo fails); RotaShield CONSISTENT
Brazil Census, maternal DTaP (pooled)	all injury admissions 0.51 (0.29-0.83)	0.77 (0.51-1.14)	DESIGN BIAS DETECTED (r < 1)
Brazil Census, maternal DTaP, SIM deaths (pooled; round two)	unintentional injury deaths 0.36 (0.15-0.79)	0.77 (0.51-1.14)	DESIGN BIAS DETECTED (r < 1)
Brazil Census, hepatitis A (report-only)	all injury admissions 2.01 (1.15-4.23)	2.17 (1.36-3.98)	DESIGN BIAS DETECTED (r > 1)
Brazil Census, hepatitis A, SIM deaths (report-only; round two)	unintentional injury deaths 0.75 (0.38-1.74)	2.17 (1.36-3.98)	MIXED

Task and design (autism context status)	Primary control: r (95% interval)	Autism on the same or own cells	Stated reading
Brazil Census, PCV10 with MenC (report-only)	all injury admissions 0.44 (0.28-0.75)	1.19 (0.76-2.05)	DESIGN BIAS DETECTED (r < 1)
Brazil Census, PCV10 with MenC, SIM deaths (report-only; round two)	unintentional injury deaths 0.26 (0.15-0.54)	1.19 (0.76-2.05)	DESIGN BIAS DETECTED (r < 1)
Brazil Census, rotavirus (report-only)	all injury admissions 0.61 (0.35-1.47)	0.95 (0.62-1.57)	NO BIAS DETECTED
Brazil Census, rotavirus, SIM deaths (report-only; round two)	unintentional injury deaths 0.37 (0.19-0.86)	0.95 (0.62-1.57)	DESIGN BIAS DETECTED (r < 1)
UK rotavirus, injury deaths (pooled contexts; different cells)	Scotland E 0.87 (0.32-2.31)	1.55 (0.38-5.31)	5 UNINFORMATIVE, 1 CONSISTENT
England rotavirus, HES and STATS19 (pooled; round two)	STATS19 casualties (S) 1.00 (0.87-1.16)	0.97 (0.89-1.06)	4 CONSISTENT, 2 UNINFORMATIVE
Scotland rotavirus, PHS and STATS19 (pooled; round two)	PHS admissions (E) 1.11 (0.42-2.75)	1.55 (0.38-5.31)	1 CONSISTENT, 3 UNINFORMATIVE
Northern Ireland rotavirus, PSNI (pooled; round two)	road casualties (S) 1.02 (0.37-2.04)	0.65 (0.29-1.32)	1 CONSISTENT, 1 UNINFORMATIVE
England MMR by local authority (report-only; round two)	admissions at ages 0-4 1.29 (0.38-4.49)	0.53 (0.36-0.96)	UNINFORMATIVE
Canada rotavirus excluding Quebec, CCDSS (pooled; round two)	asthma incidence 0.71 (0.48-1.06)	0.99 (0.32-1.89)	CONSISTENT
Norway primary care (report-only; paired outcome not autism)	injuries 1.25 (0.36-2.21)	P70-P99 0.50 (0.18-1.63)	CONSISTENT
Estonia care (report-only; paired outcome not autism)	injuries 1.87 (0.64-3.50)	F00-F99 1.08 (0.15-2.99)	CONSISTENT
Chile PCV10 discharges (report-only)	all injuries 0.95 (0.54-1.54)	0.50 (0.17-1.57)	NO BIAS DETECTED; CONSISTENT
Colombia varicella school census (pooled)	hearing 0.98 (0.66-1.49); motor or physical 0.87 (0.67-1.28); visual 0.90 (0.58-1.27)	0.75 (0.49-1.04); 0.77 (0.51-1.08)	CONSISTENT
Colombia varicella, INMLCF (pooled; round two)	transport injuries 1.36 (0.99-1.93)	0.76 (0.51-1.07)	CONSISTENT (lower limit 0.991)
Tanzania schedule (pooled)	albinism 1.00 (0.69-1.28)	1.12 (0.54-1.69)	CONSISTENT
Japan special-support classes, JE block	hearing, low vision and physical 0.91 (0.62-1.31)	1.02 (0.87-1.18)	CONSISTENT
Japan MMR, national injury deaths (pooled; different cells)	unintentional 0.86 (0.46-1.61)	0.85 (0.28-2.32)	CONSISTENT
Japan MMR, Yokohama deaths (pooled; different cells; round two)	accidents 1.00 (0.41-2.21)	0.85 (0.28-2.32)	UNINFORMATIVE
Hong Kong special education (pooled)	physical disability 1.78 (1.23-2.90)	0.89 (0.66-1.16)	BIAS SIGNAL (rule written with the estimate known)
Hong Kong injury deaths (pooled; different cells)	all injuries 0.89 (0.26-3.74)	0.89 (0.66-1.16)	UNINFORMATIVE
Hong Kong road casualties (pooled; different cells; round two)	road-traffic casualties 0.66 (0.48-0.92)	0.89 (0.66-1.16)	BIAS SIGNAL (r < 1)
Korea special education (report-only)	sensory and physical 0.90 (0.56-1.50)	0.53 (0.26-0.80)	CONSISTENT
Korea injury deaths, matched cells (report-only)	external causes 0.94 (0.25-3.50)	0.97 (0.26-3.65)	UNINFORMATIVE
Taiwan special education (report-only)	sensory and physical 0.89 (0.73-1.04)	1.07 (0.87-1.44)	NO BIAS DETECTED
Taiwan accident deaths (report-only; round two)	accident deaths 0.91 (0.48-2.05)	1.07 (0.86-1.46)	UNINFORMATIVE
Injury deaths, eight national designs	see Section S63.10	per context	5 CONSISTENT, 0 BIAS SIGNAL, 3 UNINFORMATIVE
Germany 2024 injury discharges (pooled context; cross-section only)	all injuries 0.74 (0.23-3.11)	0.87 (0.22-3.50)	UNINFORMATIVE
Sweden PCV, Socialstyrelsen patient register (pooled; round two)	all injuries 1.06 (0.98-1.13)	1.01 (0.79-1.56)	CONSISTENT

Task and design (autism context status)	Primary control: r (95% interval)	Autism on the same or own cells	Stated reading
Germany PCV, Destatis discharges on the frozen cells (pooled; round two)	all injuries 1.09 (0.31-3.91)	1.03 (0.27-3.96)	UNINFORMATIVE (prior-dominated)
Germany PCV rebuilt panel, Destatis discharges (report-only; round two)	all injuries 1.12 (0.94-1.32)	0.49 (0.21-1.08)	CONSISTENT (beside the reading)
Czechia MMR, NRHZS care register (pooled; closest cells; round two)	all injuries 1.19 (0.32-4.48)	0.91 (0.27-3.37)	UNINFORMATIVE (prior-dominated)

Notes. Readings follow each task's own stated rule, so labels differ in wording. The Chilean row carries the readings of one system under two rules (NO BIAS DETECTED under file 3335, CONSISTENT under file 3336, which counts the system once), and the two Colombian tasks share their hearing estimate, which is counted once. Round-two rows were fitted after round one under rules of their own (Section S63.1). Source: files `3326_us_core_negative_control_outcomes.md` and `3331_injury_negative_controls_us.md` through `3348_injury_negative_controls_germany_destatis.md`; for round two, `3380_injury_negative_controls_uk_round2.md`, `3381_injury_negative_controls_europe_round2.md`, `3382_injury_negative_controls_brazil_sim.md`, `3383_injury_negative_controls_us_deaths.md`, `3384_negative_controls_canada.md` and `3385_injury_negative_controls_col_hkg_jpn_twn_round2.md`.

Round two, read under each task's own rule, adds informative CONSISTENT injury controls for England and Northern Ireland rotavirus and for Sweden PCV and a CONSISTENT asthma control for Canada excluding Quebec, a second protective-direction bias signal for Brazil maternal DTaP from a second register, a protective-direction bias signal in Hong Kong road casualties and a borderline harm-direction transport-injury control for Colombia. Quebec still has no control on its own cells, the Czech controls sit on the closest cells, and the frozen German cells and the Czech cells remain untestable because every control there is prior-dominated.

The negative-control synthesis, file 3330, reads 3326, 3331-3348 and the round-two files 3380-3385 and refits nothing. It was regenerated with round two under a dated amendment that, like its original rule, was written with every per-row estimate known, and that fixed which round-two rows are carried and the new calibration sets before any pooled quantity containing them was computed. It carries 549 rows, 452 of them fitted, from 25 sources (204 rows from round two), after dropping 21 duplicate round-one rows; no source marks a round-one row as superseded, so every round-one row is carried beside round two. Every row keeps its own task's verdict; for comparison only, 3330 defines a uniform descriptive class (UNINFORMATIVE if the interval's log width is at least 0.8 x 2.7171, the prior-dominance bound, or more than 2.5 times the matching autism interval's; otherwise BIAS SIGNAL if the interval excludes one and CONSISTENT if it does not). The uniform class has no thinness criterion, so it can read a row that its own task calls thin as CONSISTENT or BIAS SIGNAL. Counted by fitted row, it gives 28 BIAS SIGNAL, 195 CONSISTENT and 229 UNINFORMATIVE readings (round one alone: 16, 114 and 142).

Counted by hierarchy context (the 15 pooled contexts and U.S. rotavirus 2006), all 16 contexts now have at least one fitted control (13 in round one), and 10 have an injury control that is informative by the uniform class (6 in round one; the new ones are Hong Kong PCV, England and Northern Ireland rotavirus and Colombia varicella). Seven contexts show a bias signal on at least one control (U.S. PCV, U.S. rotavirus 2006, Sweden PCV, Hong Kong PCV, England rotavirus, Brazil maternal DTaP and Colombia varicella; 4 in round one), 7 only CONSISTENT controls and 2 only UNINFORMATIVE ones (Quebec and Czechia MMR). Because this any-row tally grows with the number of categories a task reports, 3330 tallies the reading rows beside it: 4 contexts show a bias signal on a reading row (U.S. PCV, U.S. rotavirus 2006, Hong Kong PCV and Brazil maternal DTaP), 9 are CONSISTENT and 2 UNINFORMATIVE (U.S. RotaShield and Czechia MMR), and Quebec's only row is an analogue on other cells. The three contexts new to the any-row list enter through secondary rows: two thin Swedish fracture codes (Section S63.12), English HES poisoning admissions, 0.41 (0.19-0.85), and the Colombian transport-plus-accidental composite, 1.33 (1.01-1.96). Five fitted rows are UNINFORMATIVE by width although their interval excludes one; the only one in a hierarchy context is the U.S. PCV injury-death row, 2.59 (1.15-5.76).

3330 also previews an empirical calibration, the negative-control calibration of Schuemie and colleagues, with its method fixed in the 3330 rule before any pooled quantity was computed (the per-row estimates were already known when the rule was written). Each negative-control estimate enters as the log of its median with a standard error from its interval; a normal systematic-error distribution with mean mu and SD tau is fitted by maximum likelihood; and each autism estimate is recentred by mu and widened by tau and by the uncertainty in mu. The method assumes that systematic error is exchangeable across contexts and the same at every true effect. On the stated set of five round-one

injury controls (C_INJ, unchanged by round two), exp(mu) = 0.96 (0.51-1.87) and tau = 0.48 (0.09-1.26) on the log scale (95 percent profile-likelihood intervals). Taken at face value this systematic error is larger than the sampling error of 10 of the 16 hierarchy autism estimates (their calibrated intervals are more than twice as wide), and all 16 calibrated intervals include one, spanning 0.19 to 8.88. Tau rests on five imprecise controls, two of them from one U.S. fit, and its interval contains the assumed context-bias SDs of 0.35 and 0.70. The round-two amendment specified three sensitivity sets before computing them. S7 adds the seven informative round-two injury controls (English HES diagnoses and STATS19 casualties, Northern Irish casualties, Swedish all injuries, Brazilian SIM deaths, Hong Kong road casualties and Colombian transport injuries; n = 12) and gives exp(mu) = 0.93 (0.72-1.19), tau = 0.29 (0.00-0.62) and se(mu) = 0.11; all 16 calibrated intervals still include one, 7 are more than twice as wide, and they span 0.25 to 7.11. S7b, which also admits the U.S. and Yokohama death controls (n = 14), gives exp(mu) = 0.98 (0.76-1.26) and tau = 0.32 (0.06-0.65); S8, which keeps only the C_INJ members that pass their own placebo and harm gates (n = 3, two of them from one U.S. fit), gives exp(mu) = 1.44 (0.72-3.07) and tau = 0.28 (0.00-1.27). Across the ten specified calibration sets tau ranges from 0.15 to 0.71. The S7 interval for tau contains the smaller assumed context-bias SD, 0.35, and lies below the larger, 0.70, and its se(mu) of 0.11 is close to the shared-bias SD of 0.15 that the hierarchy assumes; S7 is a specified sensitivity, not the primary set. 3330 calls this a preview, not a result: it sizes the problem, does not replace the hierarchy and does not calibrate the shared mean. Section S62.9 carries a two-level error model fitted to the 45 same-cell controls of both rounds into the synthesis, where the calibrated common effect is 0.871 (0.741-1.020); fitted to the five C_INJ controls instead, the same construction gives 1.064 (0.711-1.564).

S63.14 What the controls show and what they do not

Designs in three places return "effects" where none can exist. In the United States the evidence is a harm-direction bias signal in the NSCH survey design and an IDEA hearing-impairment control that leans like autism under a MIXED verdict; in Brazil the Census design detects protective-direction bias for maternal DTaP in both hospital admissions and SIM deaths, for PCV10 with MenC in both and for rotavirus in deaths, and harm-direction bias for hepatitis A in admissions; in Hong Kong physical disability leans in the harm direction under a rule that was not blind, and road casualties lean in the protective direction under a rule written before any fit. Their autism estimates on those cells are not interpretable as vaccine effects. The bias runs in both directions, even within one design, which the hierarchy represents through its context-bias terms rather than a one-sided correction. The U.S. injury-death control is too imprecise to count, although its interval excludes one and its placebo fails. Elsewhere the controls are clean at their precision: English road casualties and hospital admissions, Northern Irish road casualties, Swedish patient-register injuries on the PCV block's own cells, Canadian asthma, Tanzania, Japan, Korea's special-education categories, Taiwan's special-education categories and the Swedish, German, Japanese and Chilean injury-death series. The Colombian controls are clean in round one but borderline in round two, where transport injuries lean in the harm direction against a protective autism lean and the Colombian autism estimate itself moves to 1.31 without the 2020-2021 school years; carried into the synthesis, the lockdown-free Colombian curve moves the shared mean from 0.957 to 1.045 (Section S62.8). The national Brazilian injury-death series is CONSISTENT on differently aligned cells, while SIM deaths on the autism cells are not. Agreement among designs on the same cells cannot establish specificity where the controls lean too.

Consistency is not validation. In Chile an injected doubling would still be classed CONSISTENT and the autism estimate sits near the design's no-effect value; in Tanzania, Japan and on the English road-casualty cells the autism estimate lies inside the control's interval; in Norway an injected doubling moves the control only modestly; in Taiwan the future-exposure placebo fails for controls and autism alike; in Canada the asthma control becomes a bias signal under one kernel and its prevalence sensitivity fails the placebo. Many controls are too imprecise to count: most English, Scottish and Northern Irish injury-death controls, the Scottish round-two admissions and casualties on the school-age cells, the Korean matched mortality controls, Hong Kong injury deaths, the U.S., Yokohama and Taiwanese injury deaths, the German 2024 discharges, the German discharges on the frozen block's 36 cells and the Czech care-register injuries on the closest cells. In Sweden two thin fracture codes have intervals that exclude one (lumbar or pelvic 2.02, shoulder or upper arm 1.12) but are UNINFORMATIVE under the rule because they are sparser than the autism series. Several controls are administrative placements or disability categories rather than injuries, Canadian asthma is a same-system chronic condition, national mortality series and the U.S., Hong Kong and Yokohama round-two series do not share the autism cells, the Czech round-two series sits on the closest available cells, and some controls could carry small protective pathways, such as meningitis-related hearing loss prevented by Hib or PCV, or fewer gastroenteritis visits at which asthma could first be coded after rotavirus vaccination. The rules share the three-way classification but differ in their thinness, zero-cell and prior-dominance criteria, and a few elements were fixed with partial knowledge of results: the Hong Kong rule was written with its deciding estimate known, the Chilean thin-count threshold was set with

counts in view, and several gates and sensitivities were added after the rules and are reported without changing any class. The UK, Canadian, 3381 and 3385 round-two rules were written after their data had been downloaded (for 3381, all but the Swedish external-cause series reported beside its readings), and the 3385 counts, including the thin Yokohama counts, and the 3381 reading-design counts were built before their rules. Several tasks did not preserve the exact script version that ran their stated fits and rely on hash-checked inputs, unchanged model code or bit-for-bit refits for reproduction. The controls test designs, not vaccines, and a CONSISTENT reading supplies no bound on an autism effect.

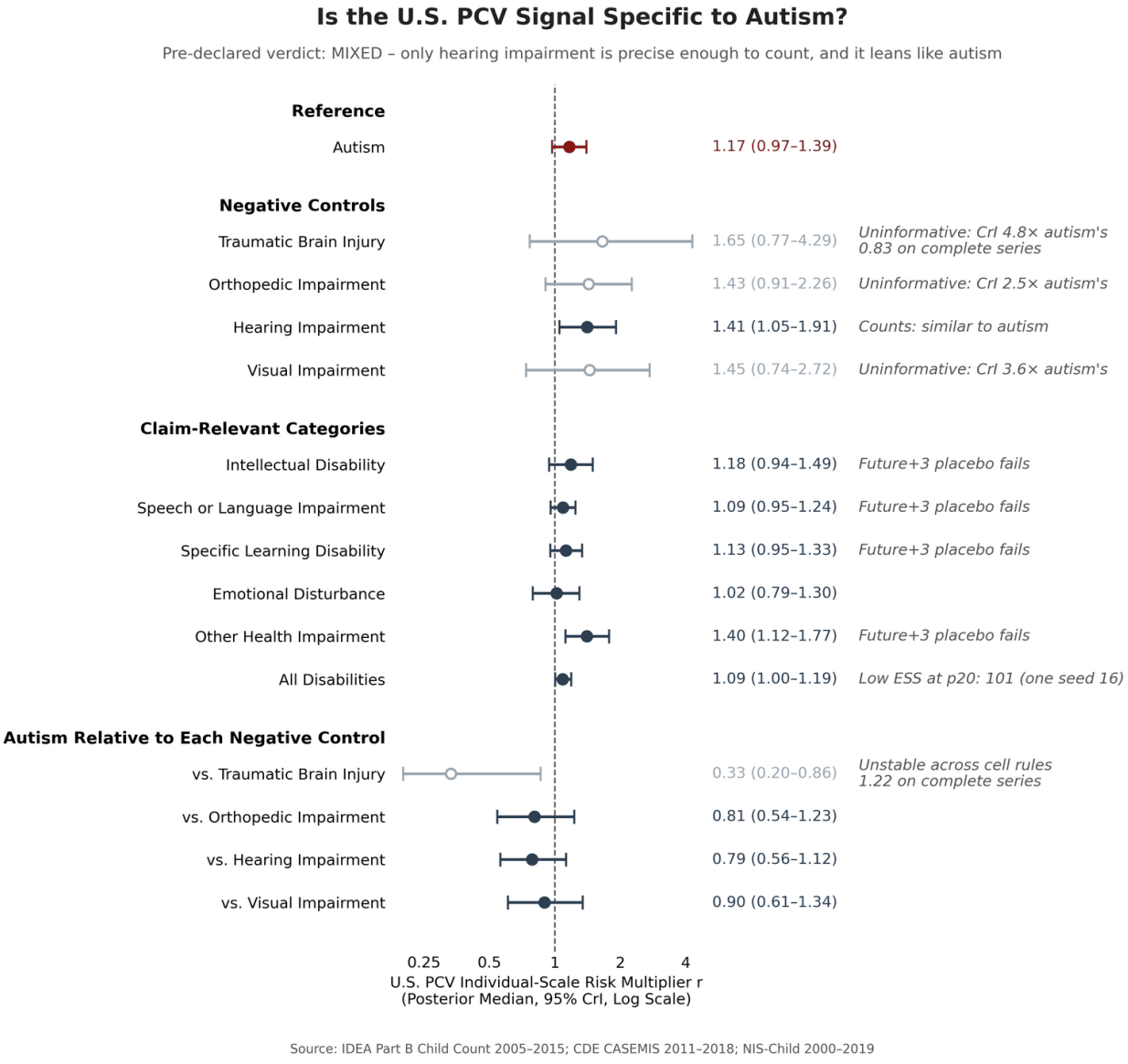


Figure S76. U.S. core PCV posterior with autism replaced by other special-education categories. U.S.-alone posterior medians and 95 percent credible intervals of the PCV r for autism, four negative-control categories, six claim-relevant categories and autism relative to each control. Hollow light markers carry no weight under the stated rule (UNINFORMATIVE or unstable across cell rules). Only hearing impairment is precise enough to count, and it leans like autism, so the verdict is MIXED.

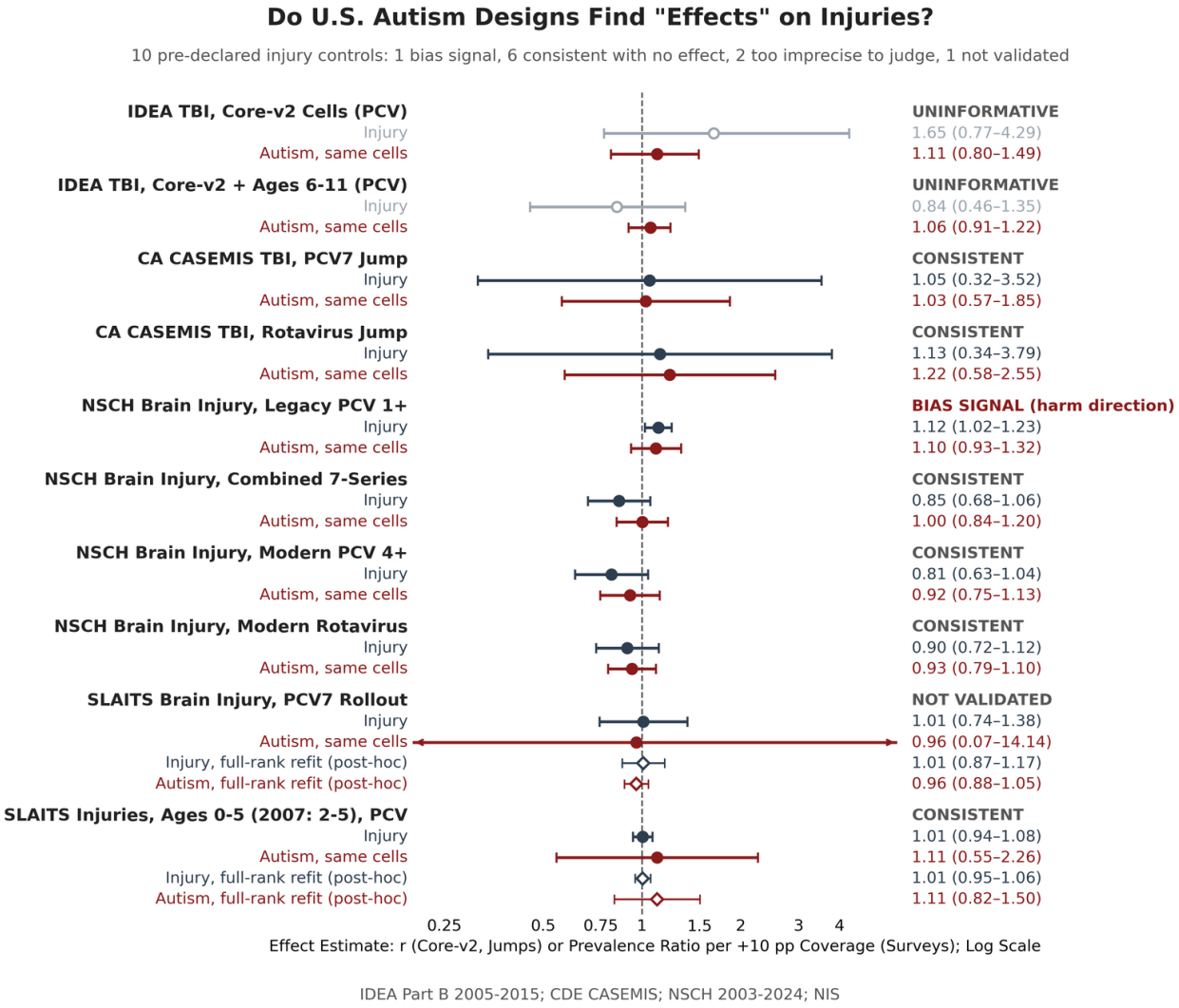


Figure S77. U.S. injury negative controls. For each of ten stated primary estimates, the injury outcome (dark) is shown beside autism on the same cells (red): IDEA traumatic brain injury in the core joint likelihood and its extension to ages 6-11, California CASEMIS traumatic brain injury for the PCV7 and rotavirus jumps, NSCH brain injury or concussion for four coverage designs, and two SLAITS designs. Hollow light markers are UNINFORMATIVE. The SLAITS rows plot post-hoc full-rank refits; under the stated rule the SLAITS PCV7-rollout row is NOT VALIDATED. Estimates are individual-scale r for the core and jump designs and prevalence ratios per ten points of coverage for the surveys.

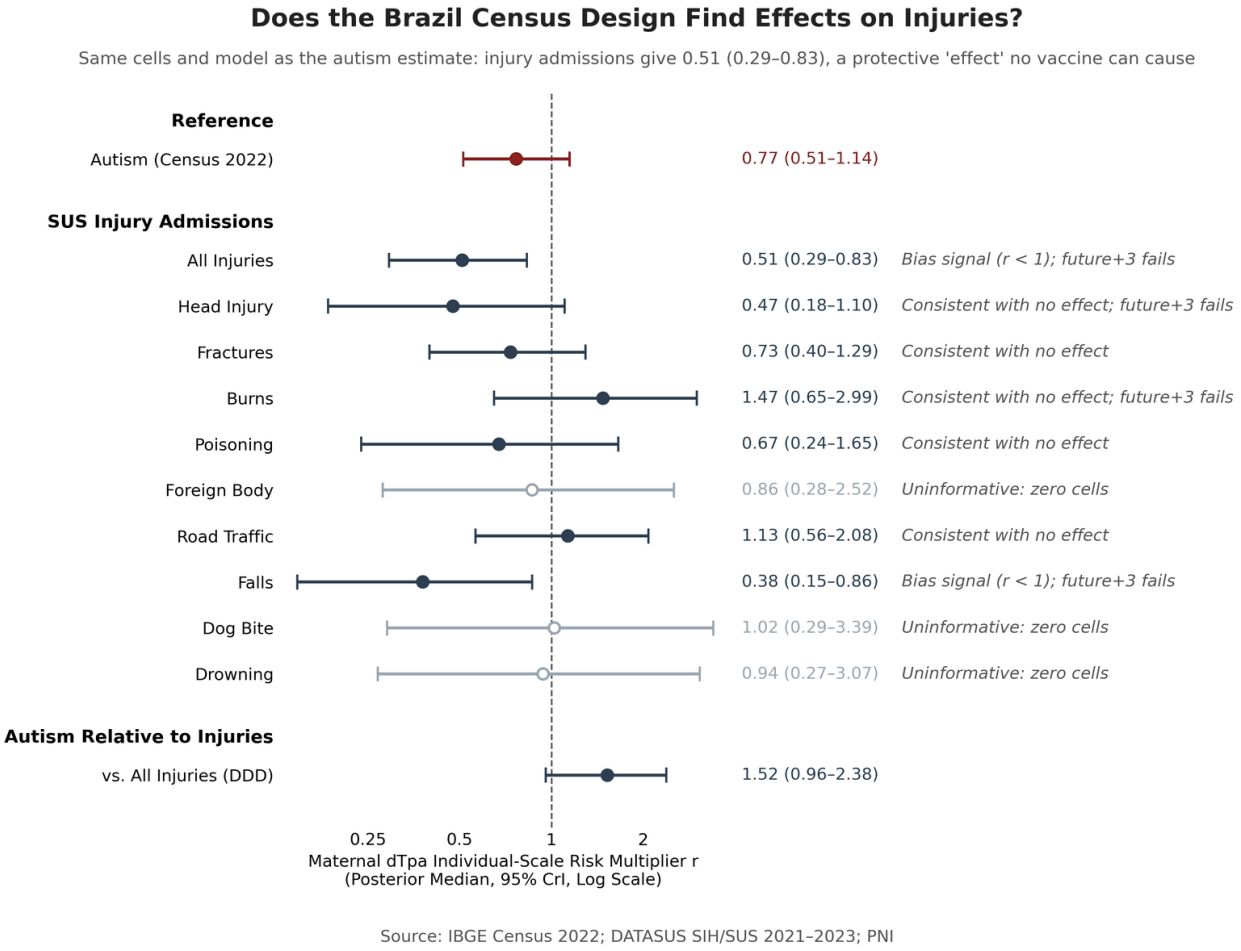


Figure S78. Brazilian Census maternal DTaP design with injury admissions. Autism, all-injury admissions, the injury components and the autism-relative (DDD) fit on the same 108 state-by-age-band cells. Hollow light markers are UNINFORMATIVE (zero-count cells or wide intervals). All injuries and falls are bias signals in the protective direction, and the stated reading is DESIGN BIAS DETECTED (r < 1).

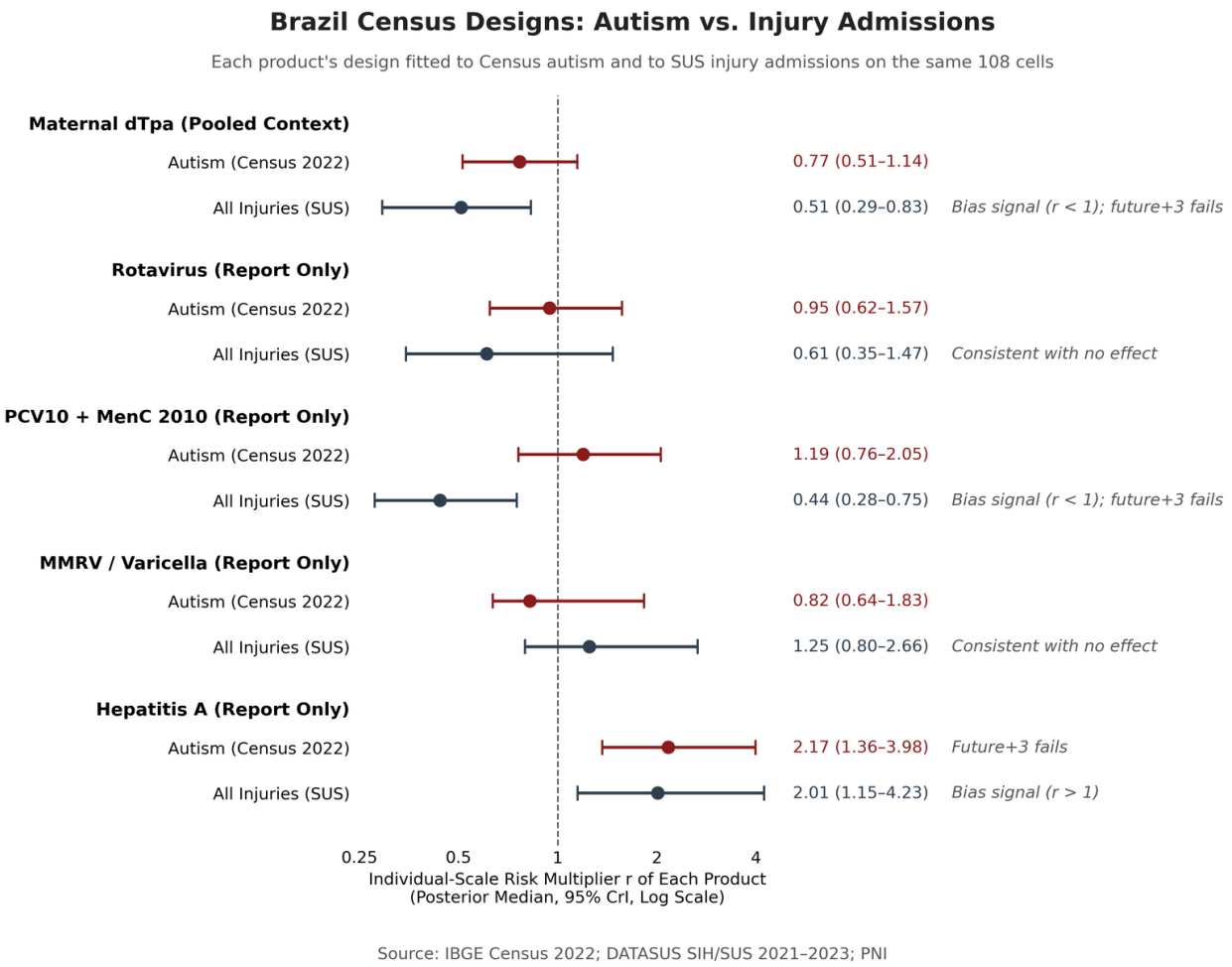


Figure S79. Brazilian Census designs by product. Autism and all-injury admissions under the maternal DTaP, rotavirus, PCV10 with MenC, MMRV and hepatitis A designs on the same cells. The five readings share cells and counts and are not independent tests.

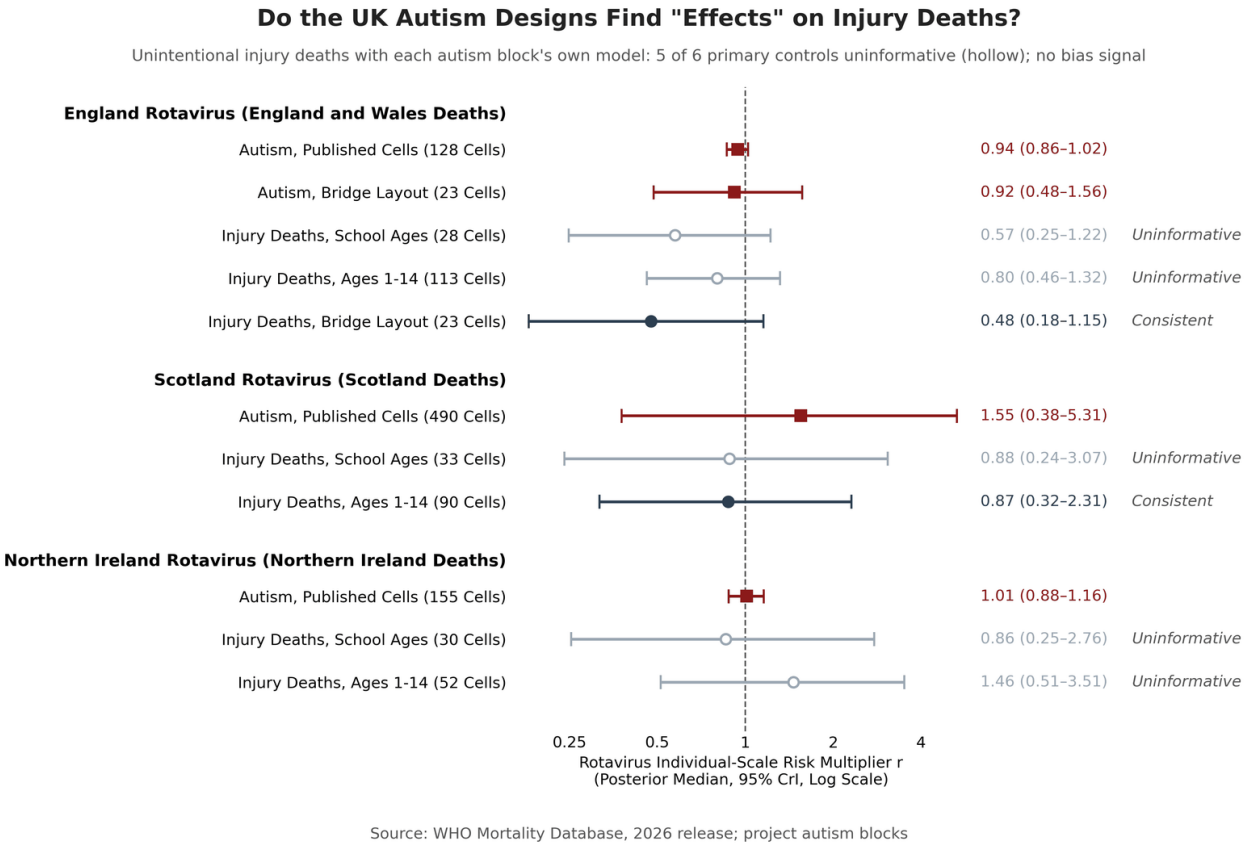
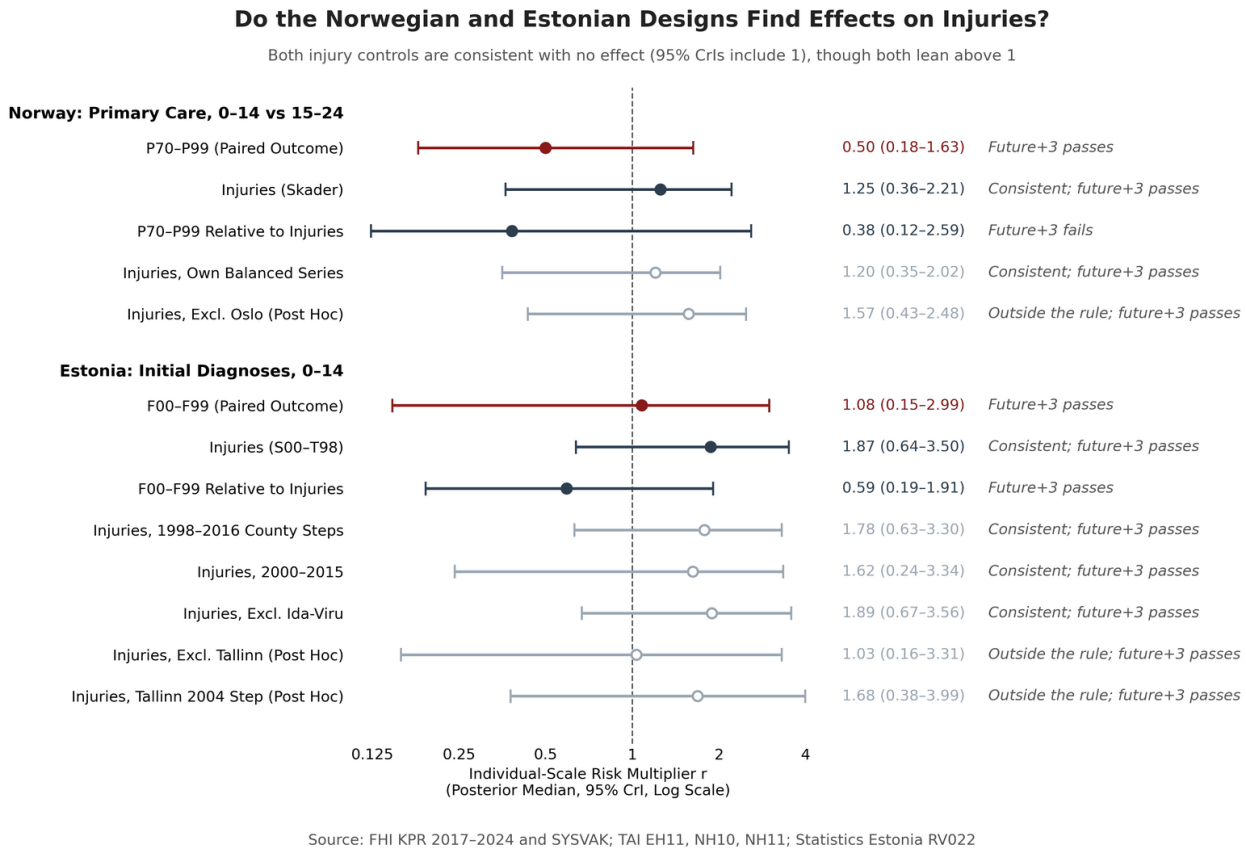


Figure S80. UK rotavirus designs with WHO Mortality Database injury deaths. For England, Scotland and Northern Ireland the published autism posterior is shown beside unintentional injury deaths on the cell set closest to the autism cells and on an early-age extension, with the 23-cell English bridge on which autism and injury share one layout. Hollow markers are UNINFORMATIVE; five of six primary controls are uninformative and none is a bias signal.



Source: FHI KPR 2017–2024 and SYSVAK; TAI EH11, NH10, NH11; Statistics Estonia RV022

Figure S81. Norwegian and Estonian injury controls. Injury care contacts replace the paired report-only outcomes, Norwegian symptom codes P70-P99 and Estonian F00-F99 care, which are not autism, on the same cells, with DDD fits and sensitivities. Hollow grey markers are sensitivity and post-hoc rows. Both injury controls are CONSISTENT, and both lean above one.

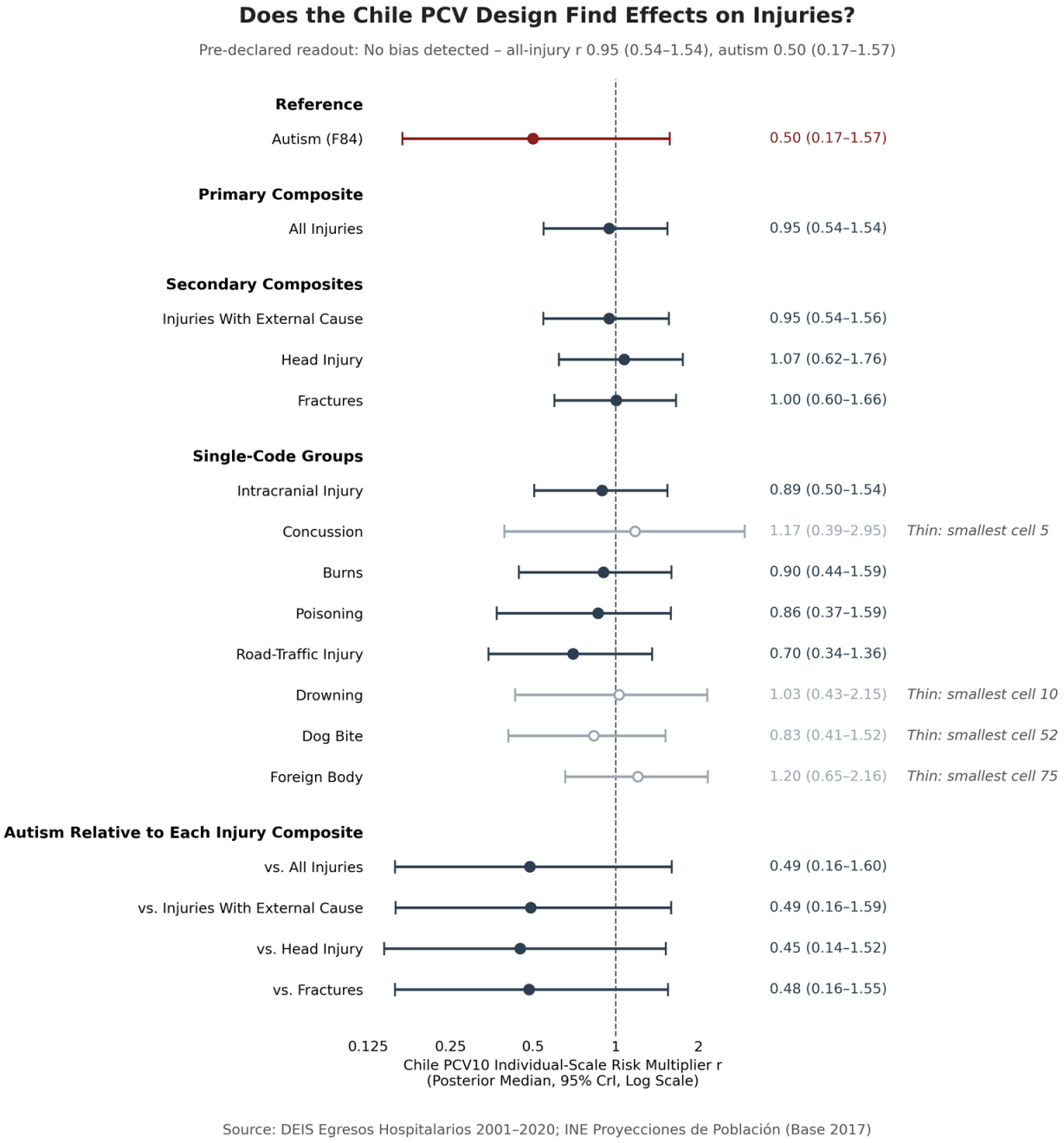


Figure S82. Chilean PCV10 discharge design with injury discharges, first rule. Autism (F84), the all-injury composite, three secondary composites, eight single-code groups and DDD fits on the same 40 national cells. Hollow light markers are UNINFORMATIVE under the rule. The stated readout is NO BIAS DETECTED, although an injected doubling would still be classed CONSISTENT.

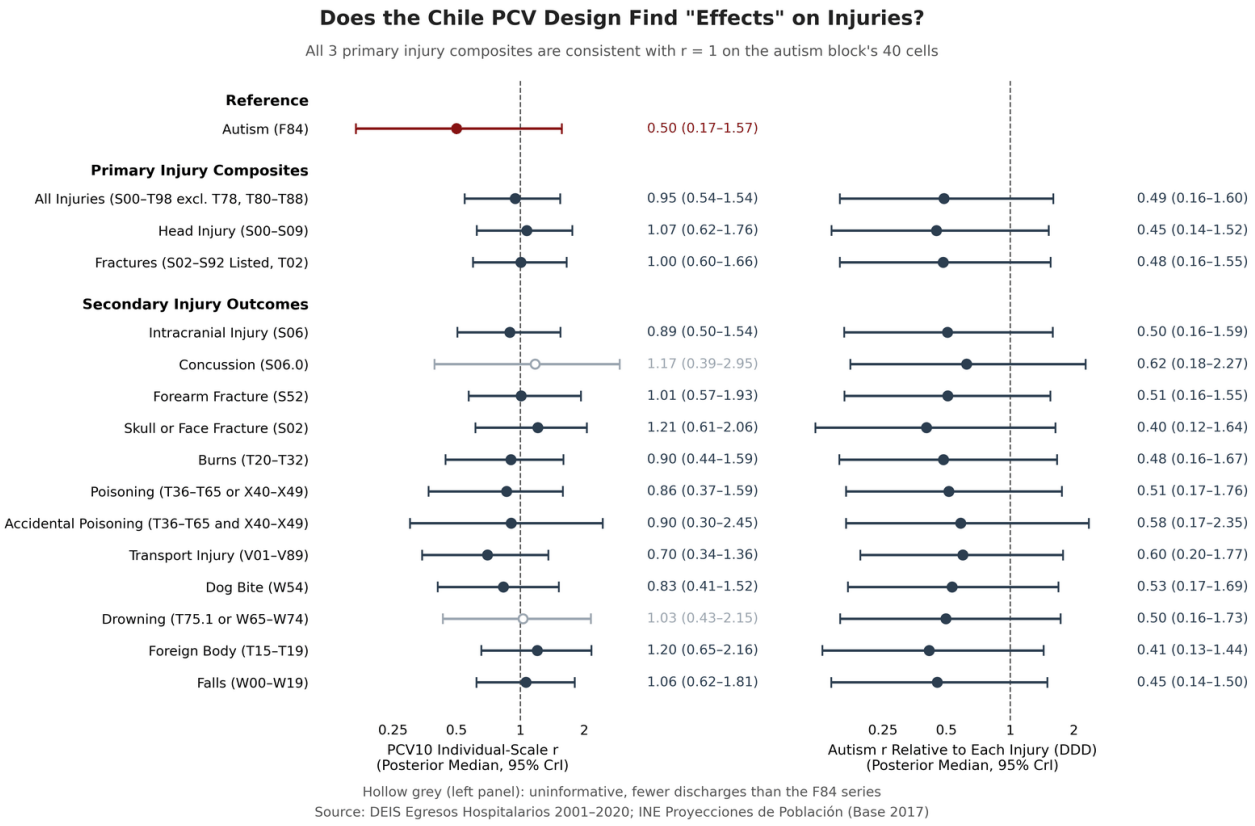
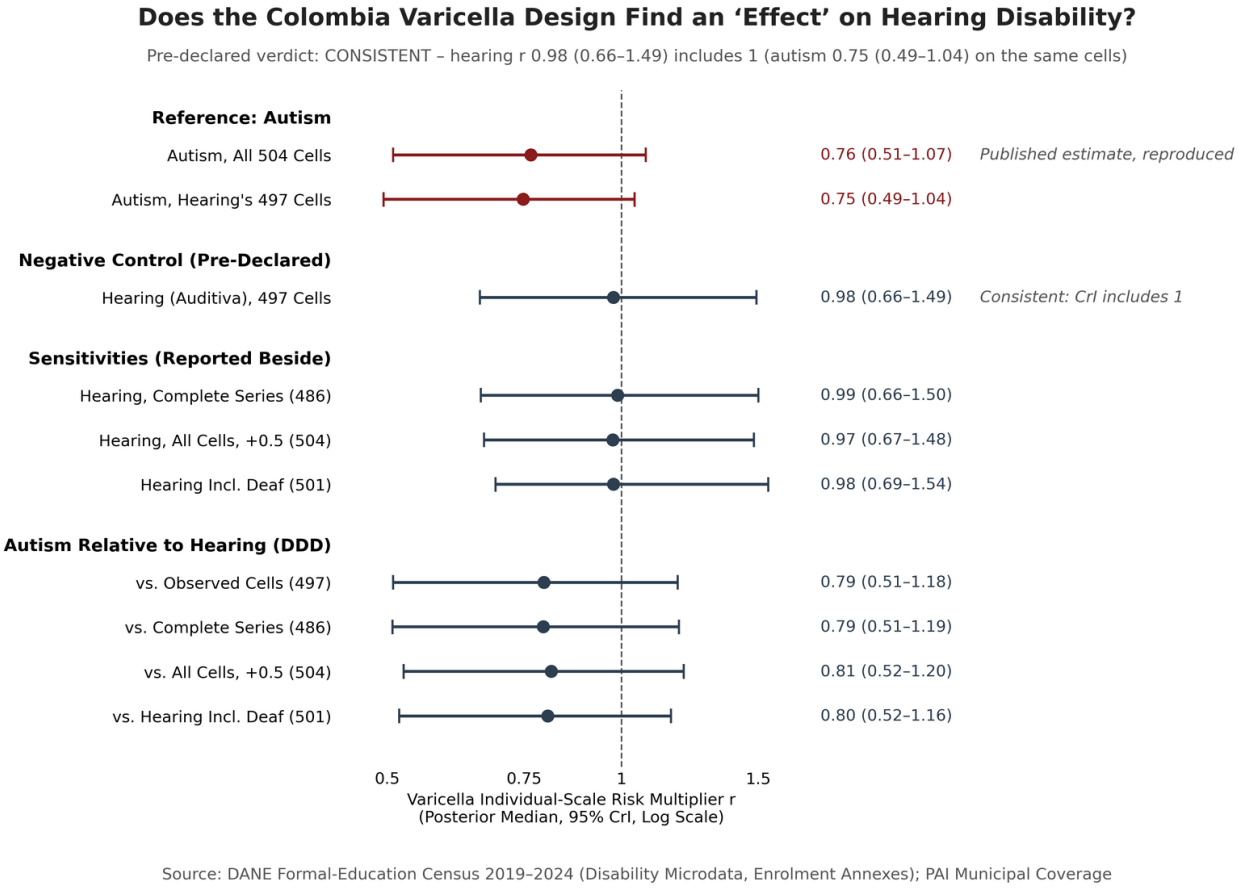


Figure S83. Chilean PCV10 discharge design, second rule. All 15 injury outcomes (left) and autism-relative DDD fits (right) on the same 40 cells; hollow grey markers are outcomes thinner than the F84 series. The three primary composites are CONSISTENT, and the autism median lies below each of their lower limits.



Source: DANE Formal-Education Census 2019–2024 (Disability Microdata, Enrolment Annexes); PAI Municipal Coverage

Figure S84. Colombian varicella school-census design with hearing disability. Autism on all 504 cells and on the 497 hearing cells, the primary hearing control, three sensitivities and DDD fits. Hearing disability is CONSISTENT.

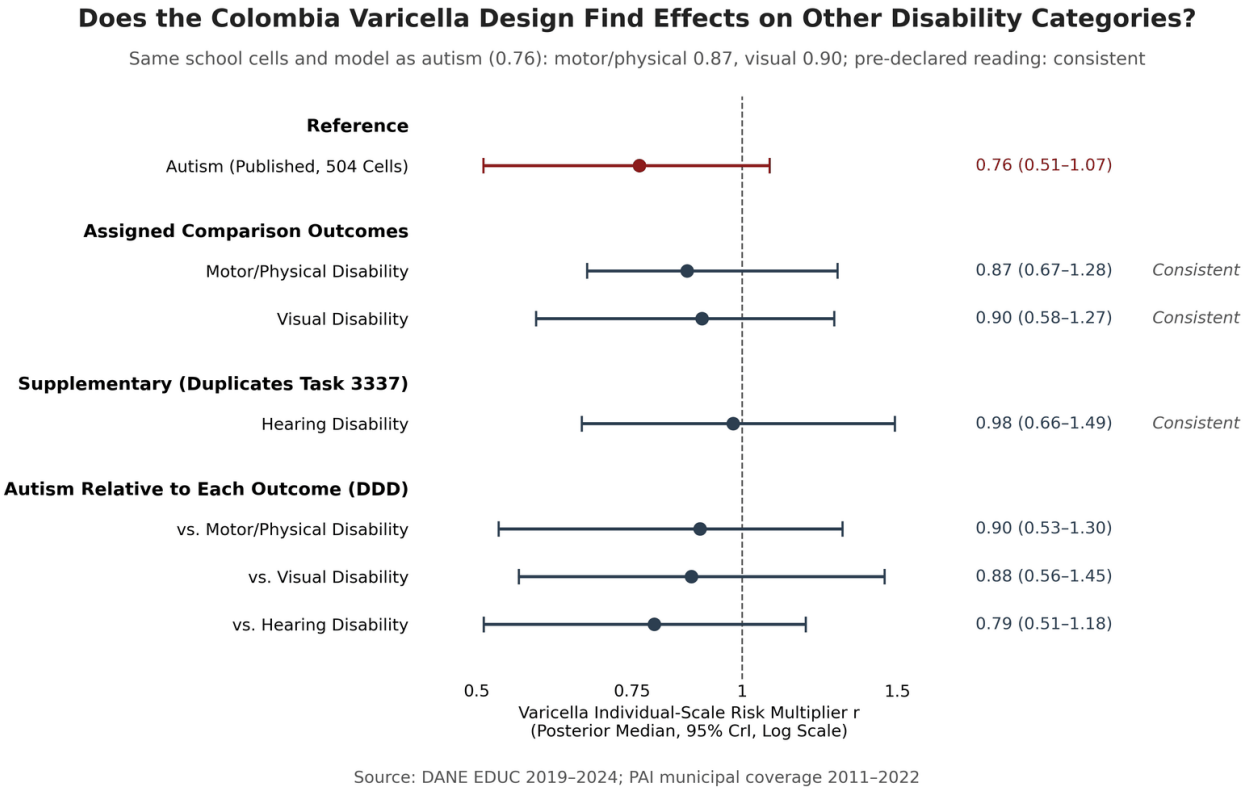


Figure S85. Colombian varicella design with motor or physical and visual disability. Autism, the two assigned specificity outcomes, hearing (the estimate of Figure S84) and DDD fits on the same school-census cells. Both assigned outcomes are CONSISTENT; the census has no injury category.

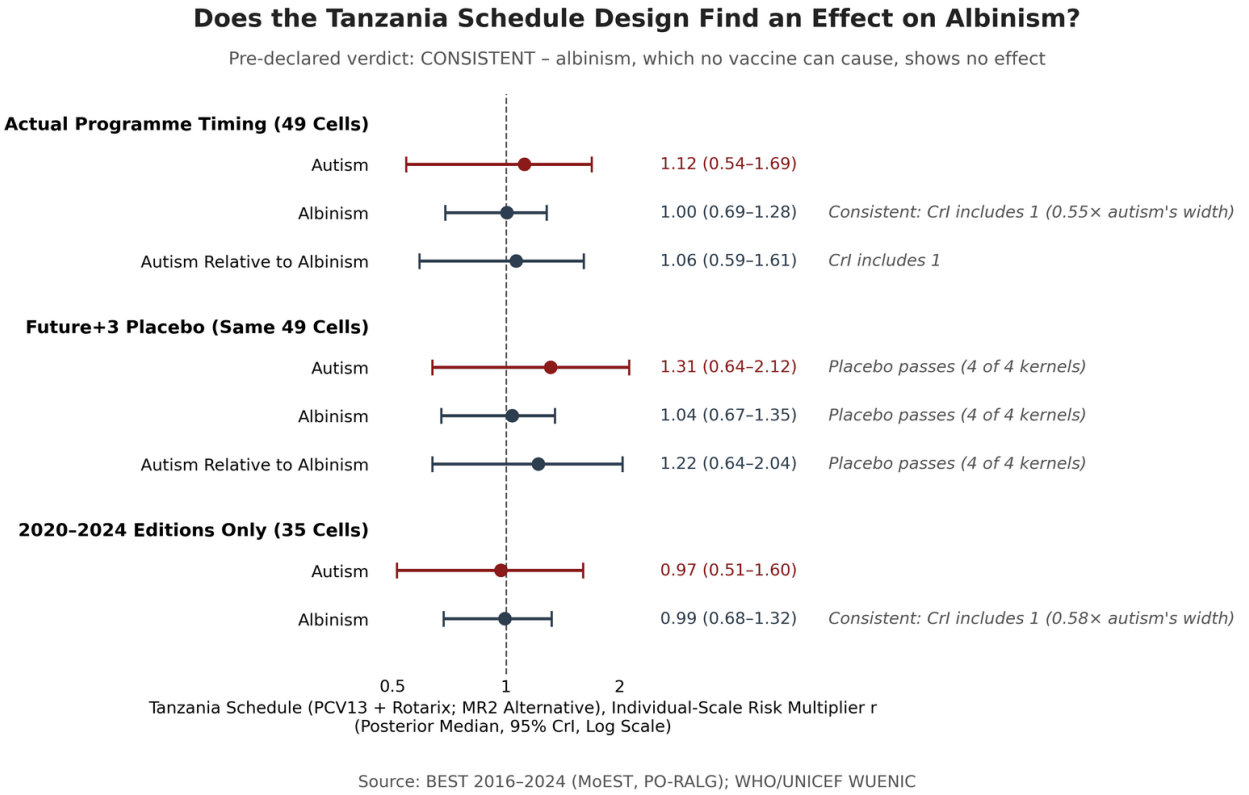


Figure S86. Tanzanian schedule design with albinism. Autism, albinism and the DDD under the actual timing, the future+3 placebo and the 2020-2024 edition sensitivity. Albinism is CONSISTENT, but the autism median lies inside its interval.

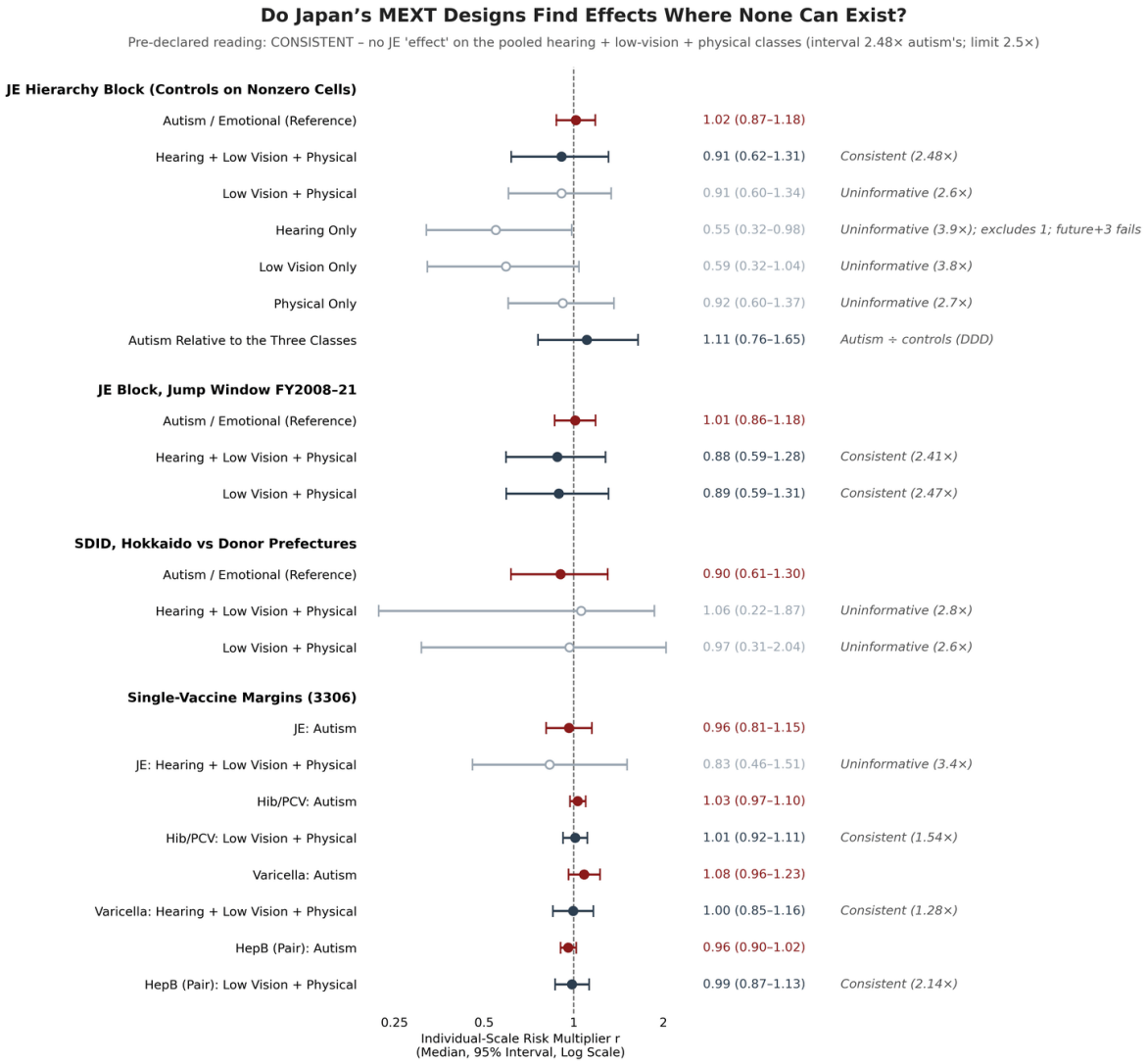


Figure S87. Japanese special-support class controls. Autism and emotional-disorder classes (reference) beside hearing, low-vision and physical classes in the Japanese encephalitis hierarchy block, its jump window, the Hokkaido synthetic control and the single-vaccine margins. Multipliers give each interval's width relative to autism's on the same cells; hollow markers exceed the 2.5-fold limit. The headline reading is CONSISTENT at 2.48 times autism's width.

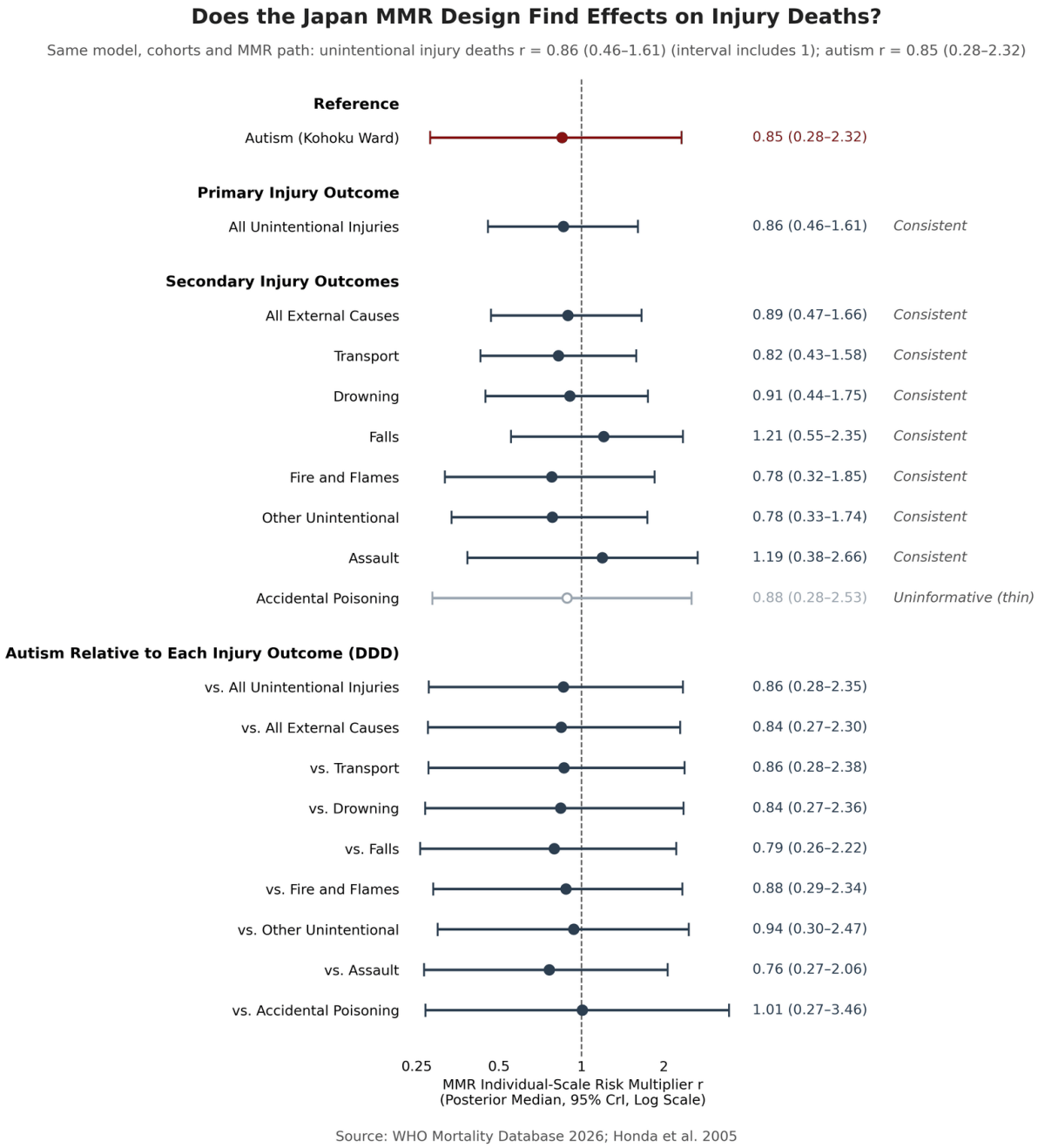
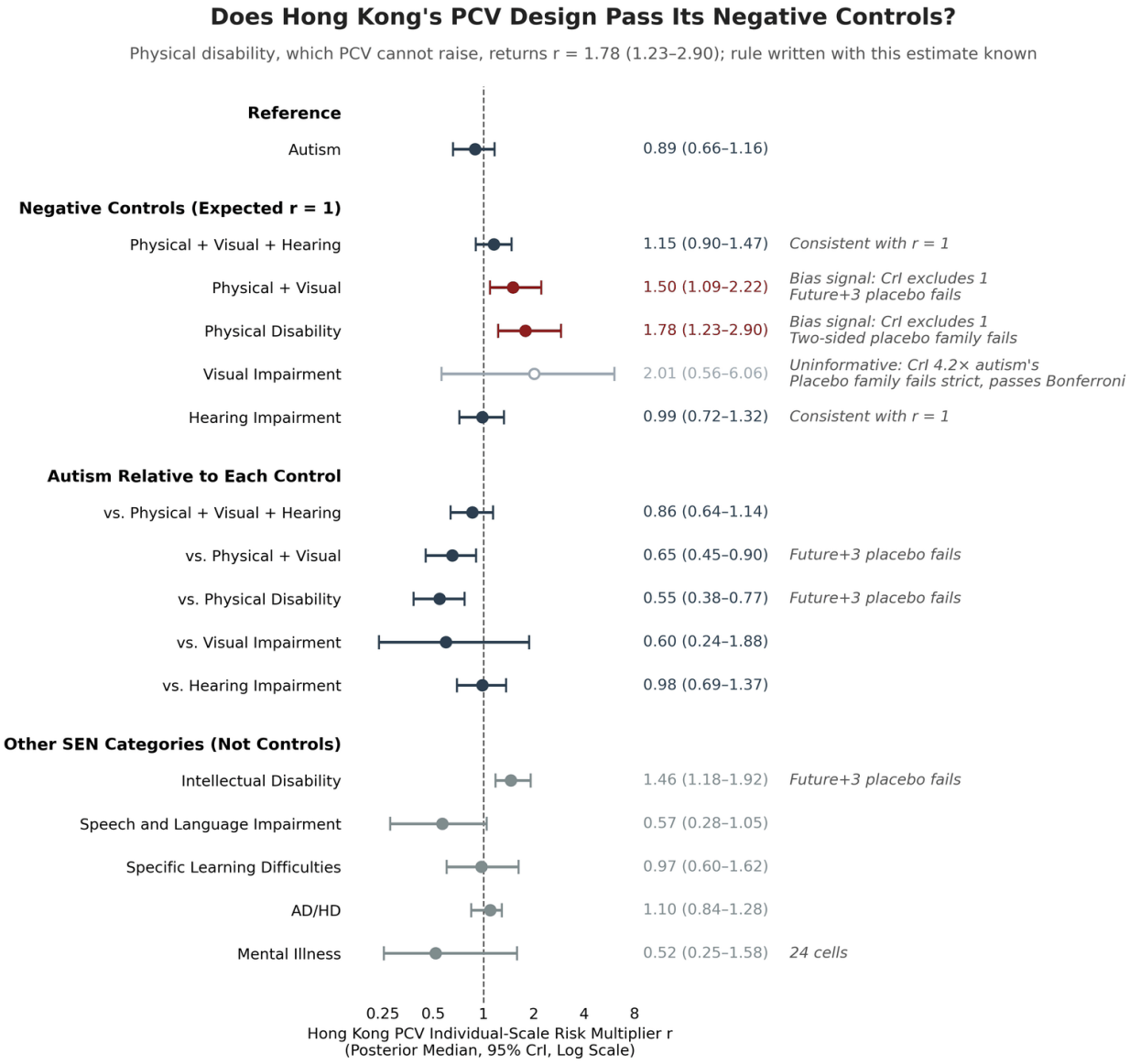


Figure S88. Japan MMR design with national injury deaths. Autism in Kohoku ward beside national unintentional injury deaths, eight secondary injury outcomes and DDD fits on the same nine birth cohorts and MMR coverage path. Hollow light markers are UNINFORMATIVE. The primary control is CONSISTENT and sits where the autism estimate does.



Source: EDB SEN pupils by grade, 2015/16–2019/20 (LegCo replies 2021–22); EDB enrolment 2015–2019; CHP immunisation survey 2015

Figure S89. Hong Kong PCV design with other special-education categories. Autism, five negative controls and other categories on the same 60 grade-by-year cells, with DDD fits. Red marks bias signals, hollow light markers UNINFORMATIVE controls and grey categories outside the rule. Physical disability and physical plus visual impairment are bias signals in the harm direction; the rule was written with the deciding estimates known.

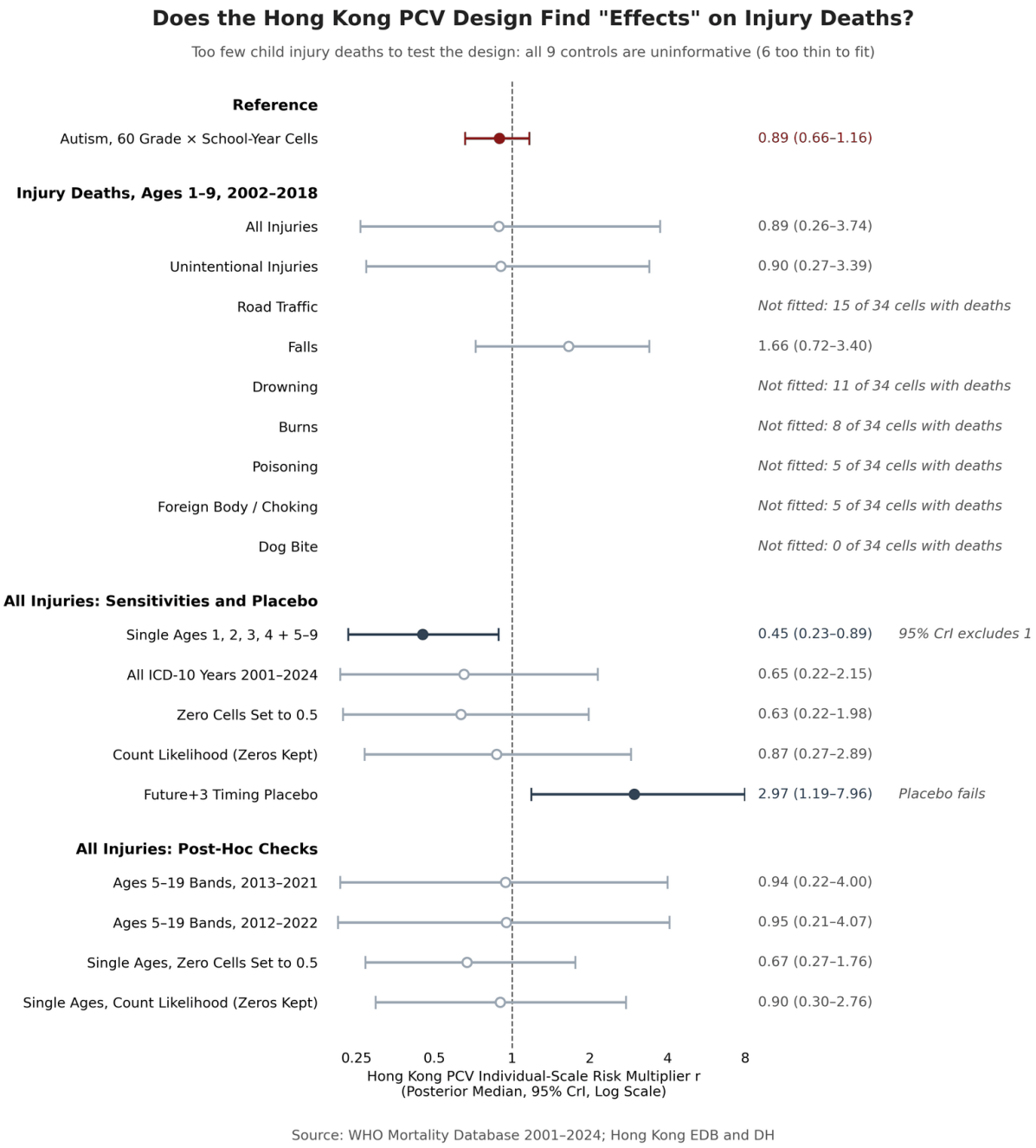


Figure S90. Hong Kong PCV design with injury deaths. Autism on its 60 cells beside injury deaths at ages 1-9 in 2002-2018, sensitivities, the future+3 placebo and post-hoc checks. Too few deaths exist to test the design: every control is UNINFORMATIVE, and six are too thin to fit.

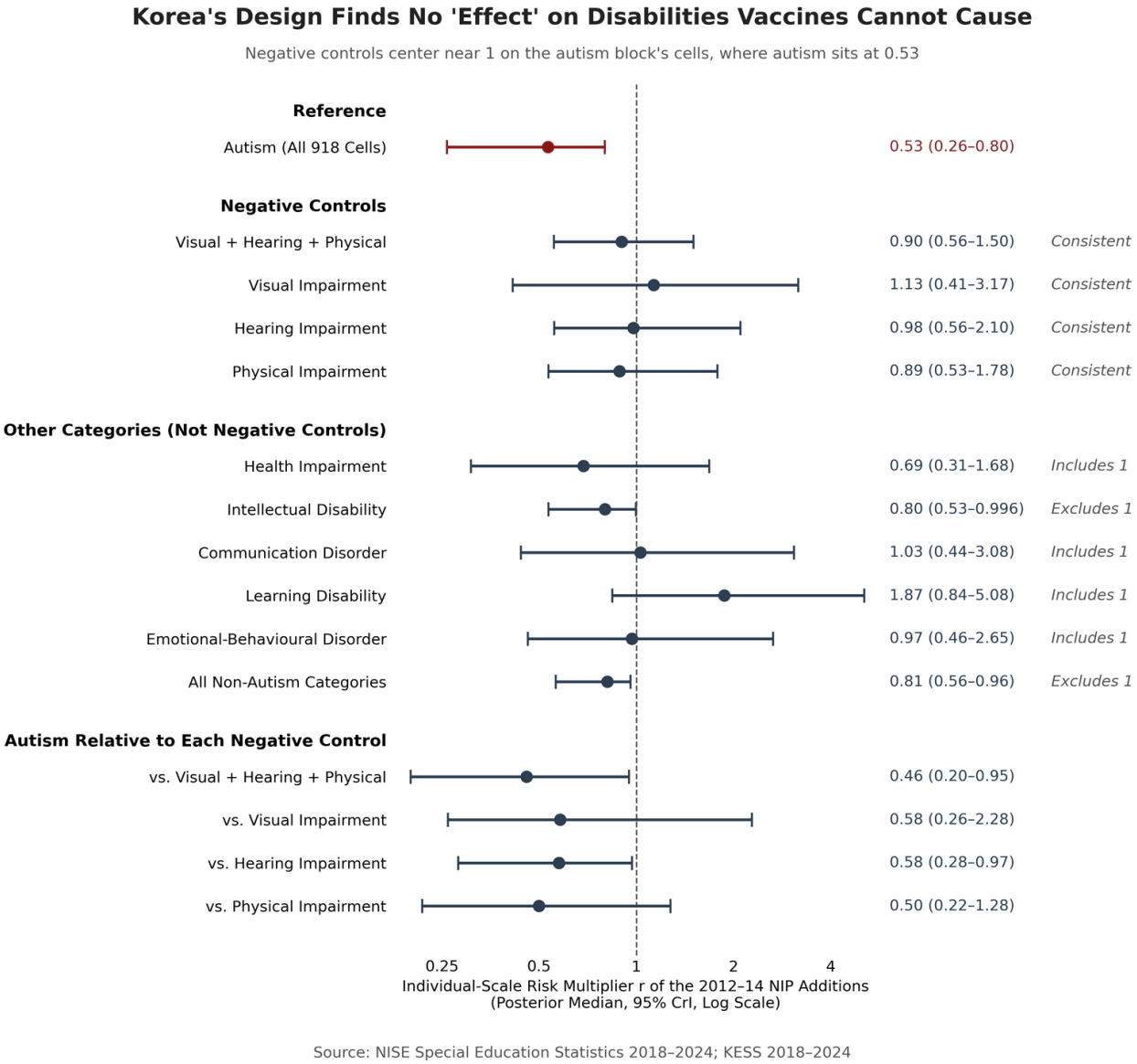


Figure S91. Korean special-education design with sensory and physical categories. Autism on all 918 cells beside four negative controls, other categories that are not negative controls, and autism relative to each control. All four controls are CONSISTENT and centre near one, while autism sits at 0.53.

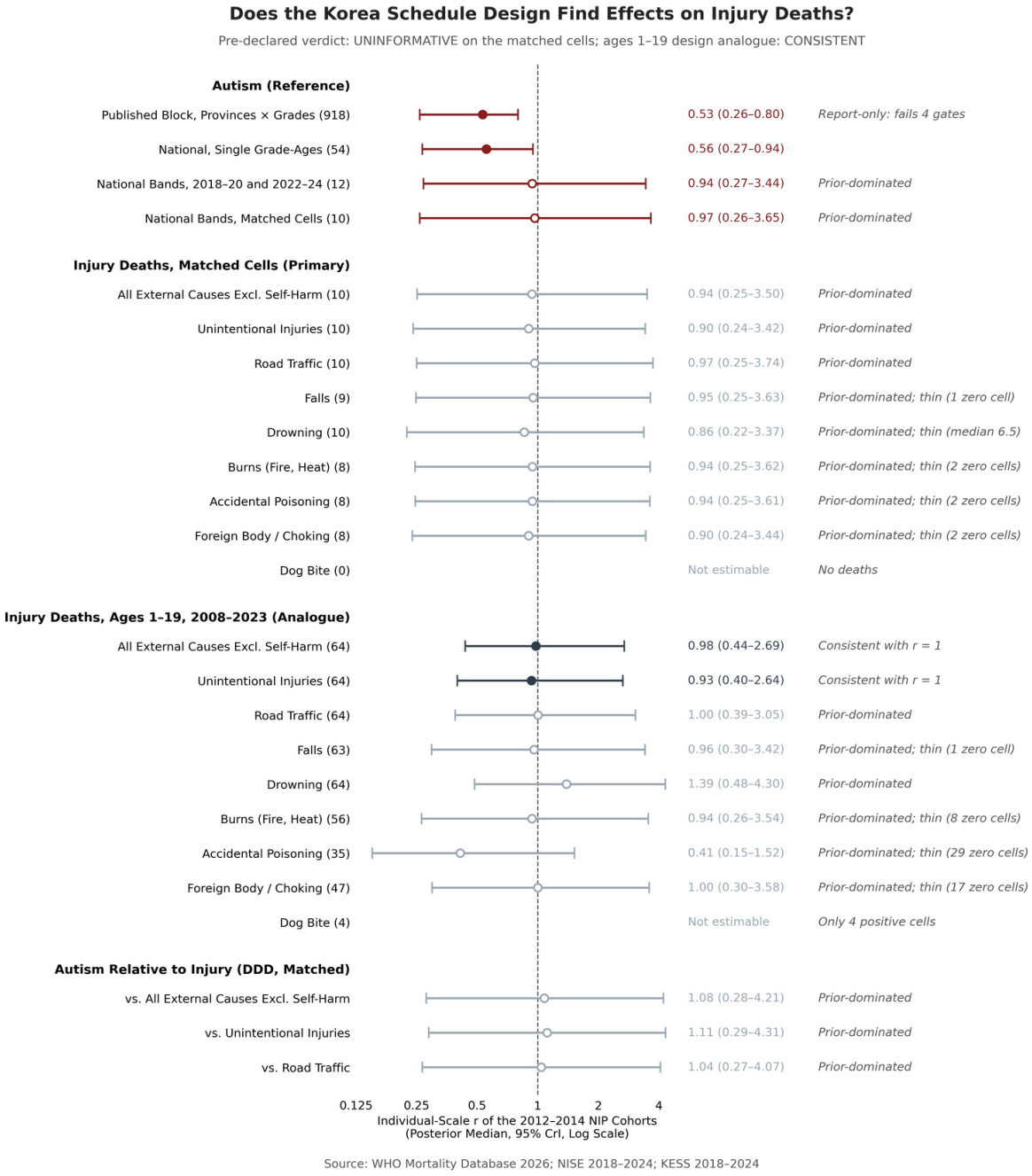


Figure S92. Korean schedule design with WHO Mortality Database injury deaths. Injury outcomes on the cells matched to the autism block, on a secondary design over ages 1-19 in 2008-2023 and on a 1997-2023 sensitivity. Hollow light markers are UNINFORMATIVE under the rule. The matched cells are prior-dominated; the secondary design reads CONSISTENT.

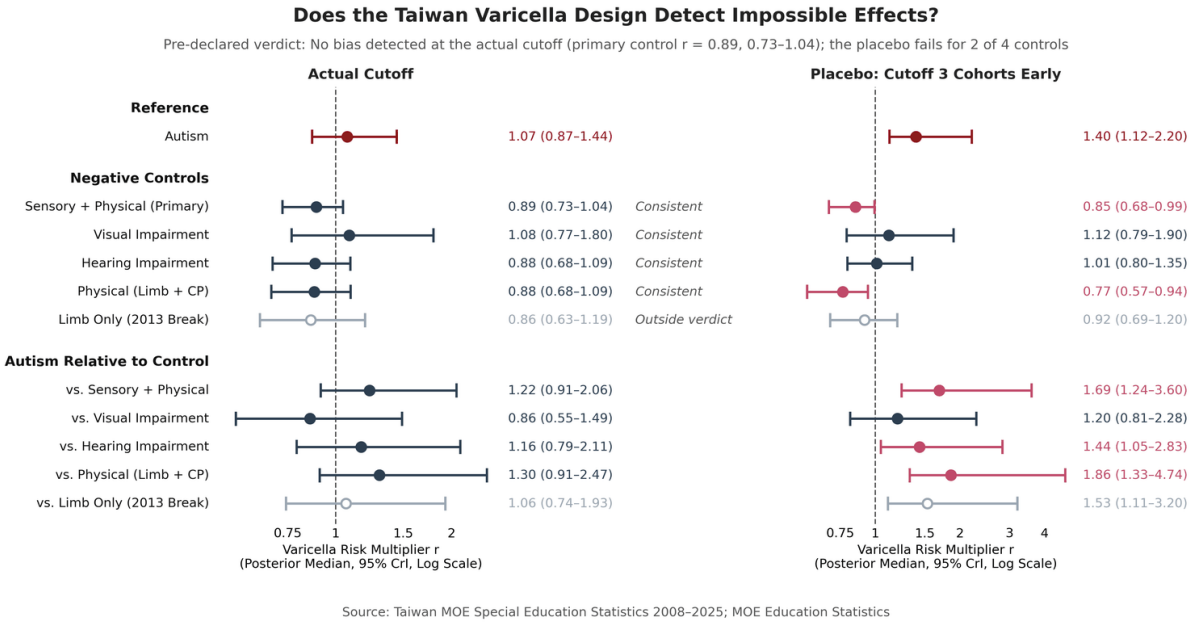


Figure S93. Taiwanese varicella design with sensory and physical special-education categories. Autism, four negative controls, the limb-only series and DDD fits at the actual cutoff (left) and with the cutoff three cohorts early (right). The primary control is CONSISTENT, but the placebo fails for two of four controls, as it does for autism.

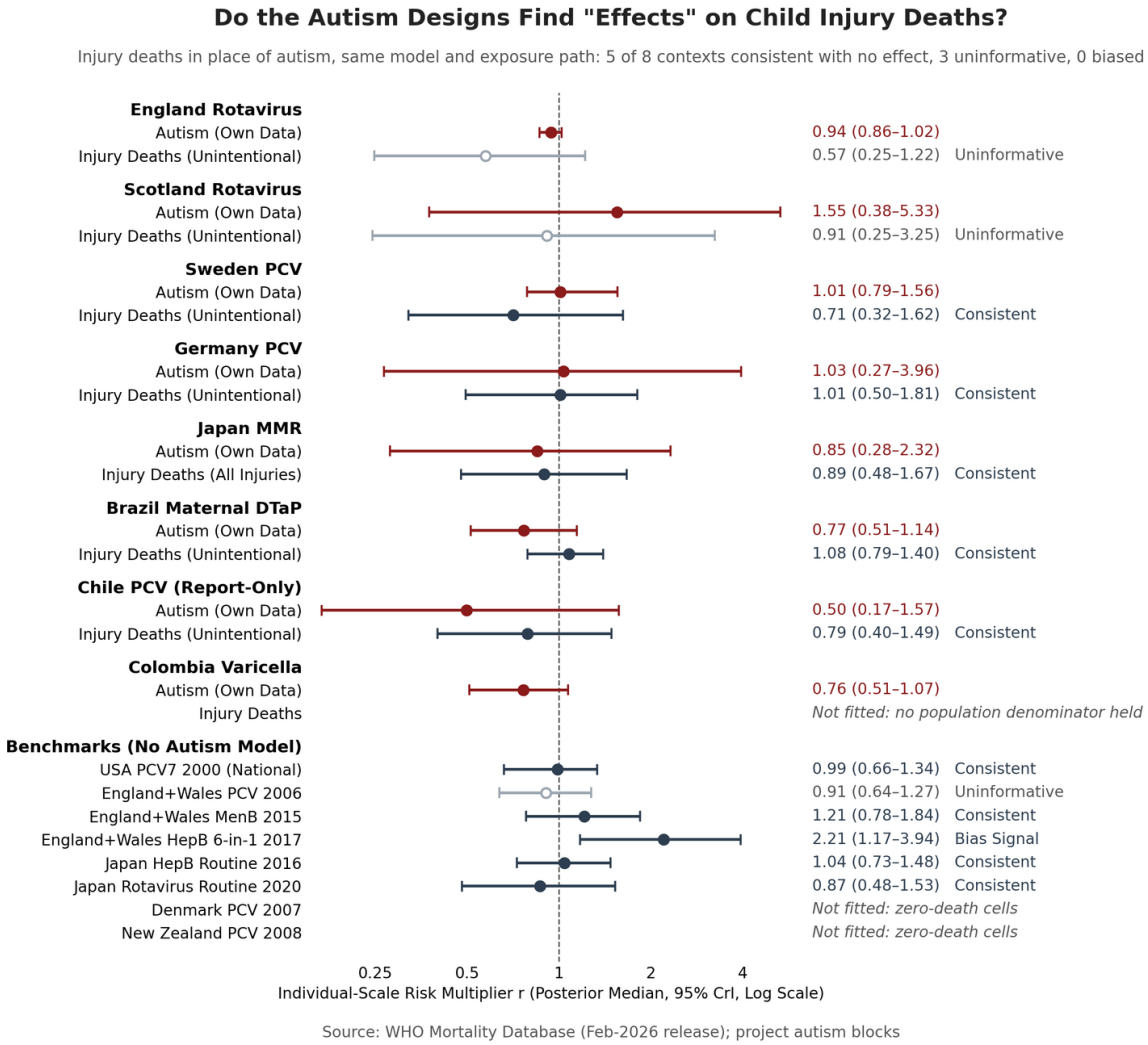


Figure S94. Injury deaths across eight national autism designs. For each context the autism posterior on its own data (red) is shown beside WHO Mortality Database injury deaths under the same model and exposure path, with benchmark introductions that have no autism model. Hollow light markers are UNINFORMATIVE. Five contexts are CONSISTENT, three UNINFORMATIVE and none a bias signal; the English hexavalent hepatitis B benchmark is a bias signal on cells that include the pandemic years.

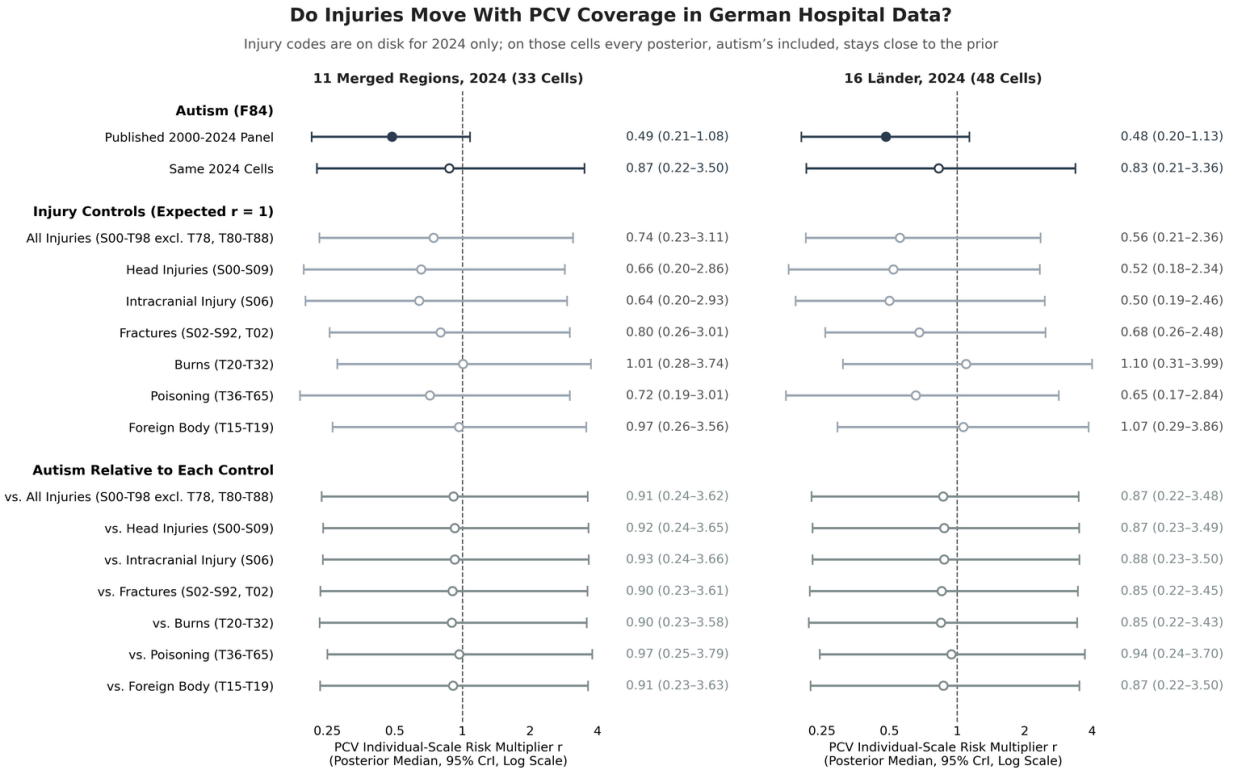


Figure S95. German PCV design with 2024 Destatis injury discharges. Autism, seven primary injury controls, DDD fits and single-code controls in the merged-region and 16-Land designs. Red marks a bias signal, light hollow markers injury controls classed UNINFORMATIVE, hollow slate and grey markers prior-dominated autism and DDD posteriors, and grey the DDD fits. The 2024 cross-section contains only post-programme cohorts, and every primary control is UNINFORMATIVE.

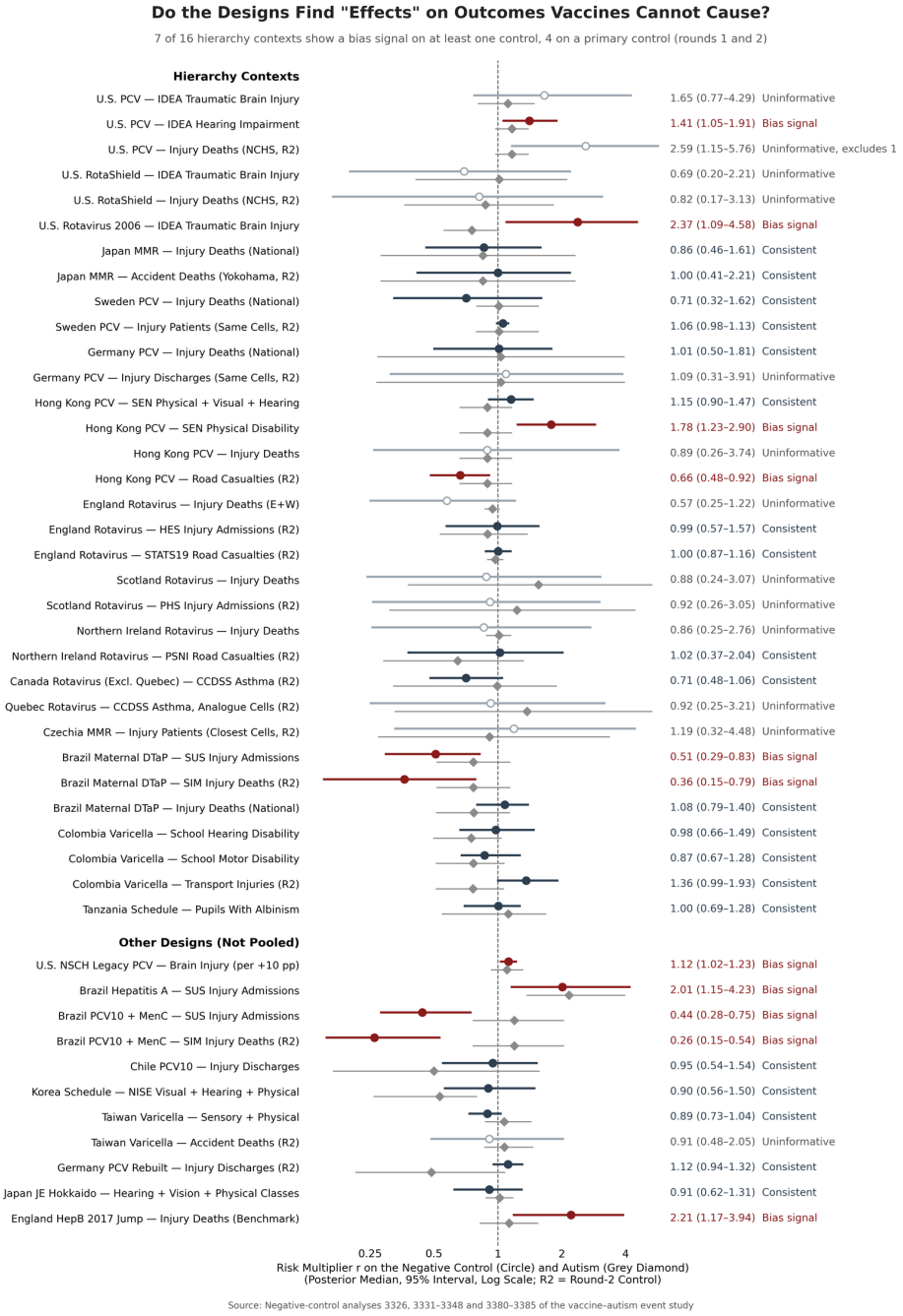


Figure S96. Negative controls by context, rounds one and two (file 3330). For each hierarchy context the reading controls (circles; R2 marks a round-two control) are shown beside autism on the same or own cells (grey diamonds), followed by designs outside the pooled hierarchy. Red marks a bias signal and hollow light markers an UNINFORMATIVE control under the uniform descriptive class of file 3330, which has no thinness criterion (the thin Yokohama row reads CONSISTENT here); each row keeps its own task's verdict in the text. Seven of the 16 hierarchy contexts show a bias signal on at least one control and four on a reading control; the secondary rows that add Sweden, England and Colombia to the seven are not drawn. The U.S. PCV injury-death row is UNINFORMATIVE although its interval excludes one.

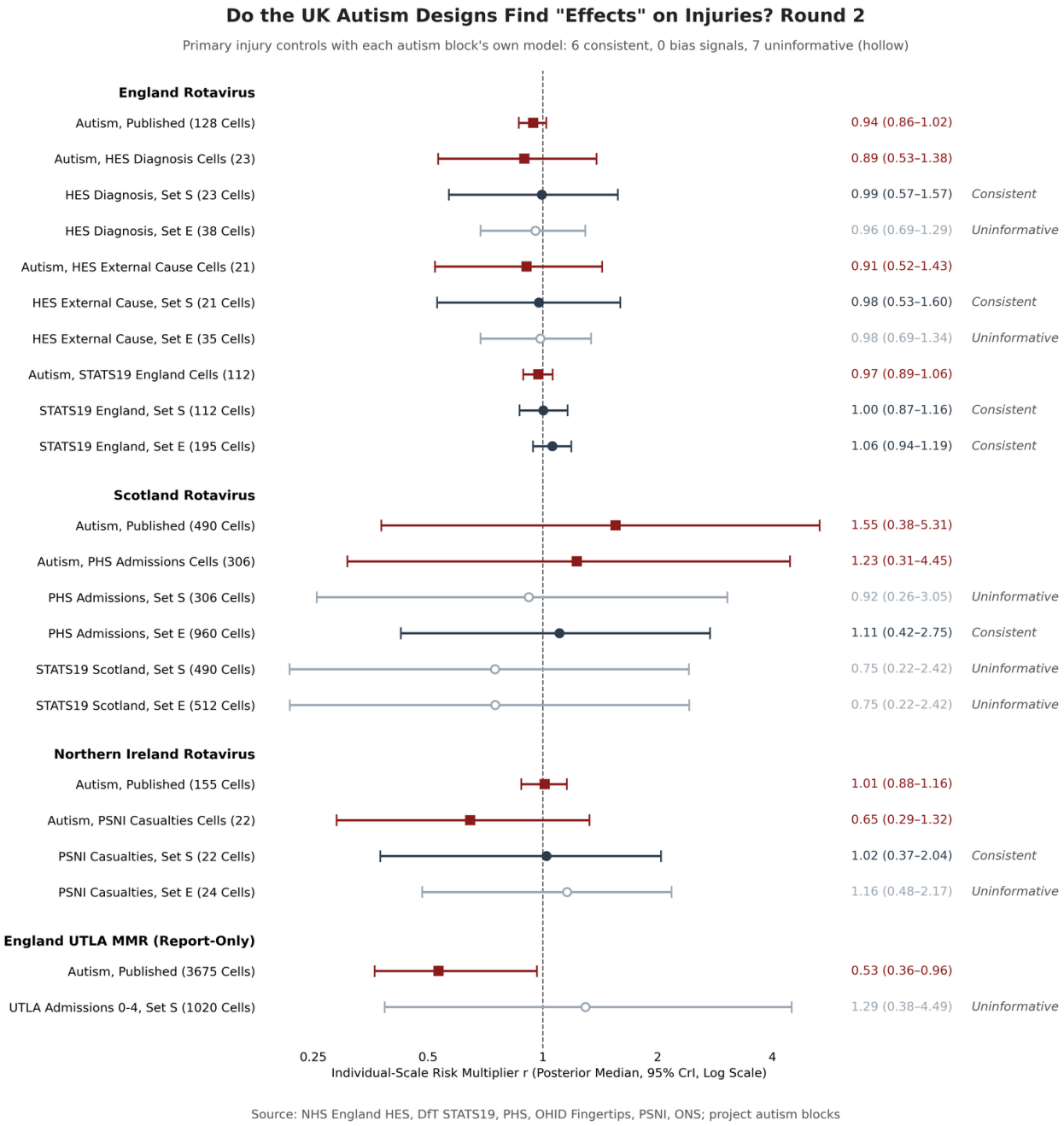


Figure S97. UK autism designs with round-two injury series. For England, Scotland and Northern Ireland rotavirus the published autism posterior and autism refitted on each control's cells (red) are shown beside HES admissions by primary diagnosis and by external cause, STATS19 and PSNI road casualties and PHS admissions on the same-layout set S and the extension E, followed by the report-only English MMR design with injury admissions at ages 0-4. Hollow light markers are UNINFORMATIVE. Six primary controls are CONSISTENT, seven UNINFORMATIVE and none a bias signal.

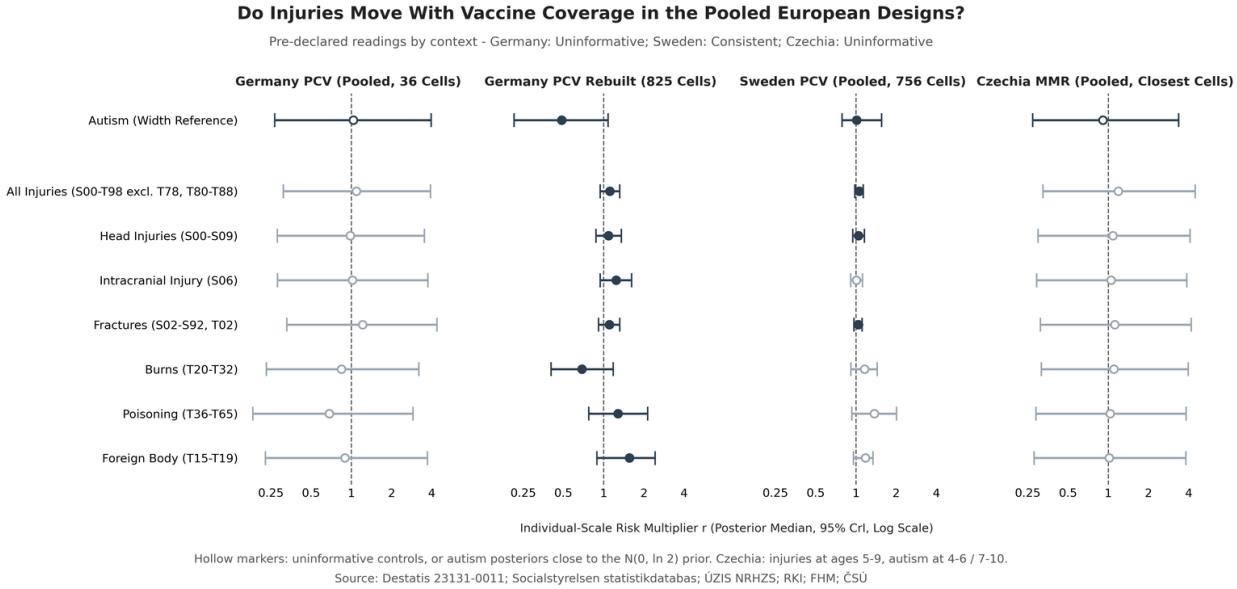


Figure S106. Pooled European designs with round-two injury controls. For the frozen German PCV block (36 cells), the report-only rebuilt German panel (825 cells), the Swedish PCV block (756 cells) and the Czech MMR model, autism refitted as the width reference (top) is shown beside the seven primary injury controls; the Czech controls sit on the closest available cells (injuries at ages 5-9, autism at 4-6 and 7-10). Hollow markers are UNINFORMATIVE controls or autism posteriors close to the N(0, ln 2) prior. Sweden reads CONSISTENT, all injuries 1.06 (0.98-1.13) against autism 1.01 (0.79-1.56) on the same cells; the frozen German and Czech readings are UNINFORMATIVE because every control is prior-dominated.

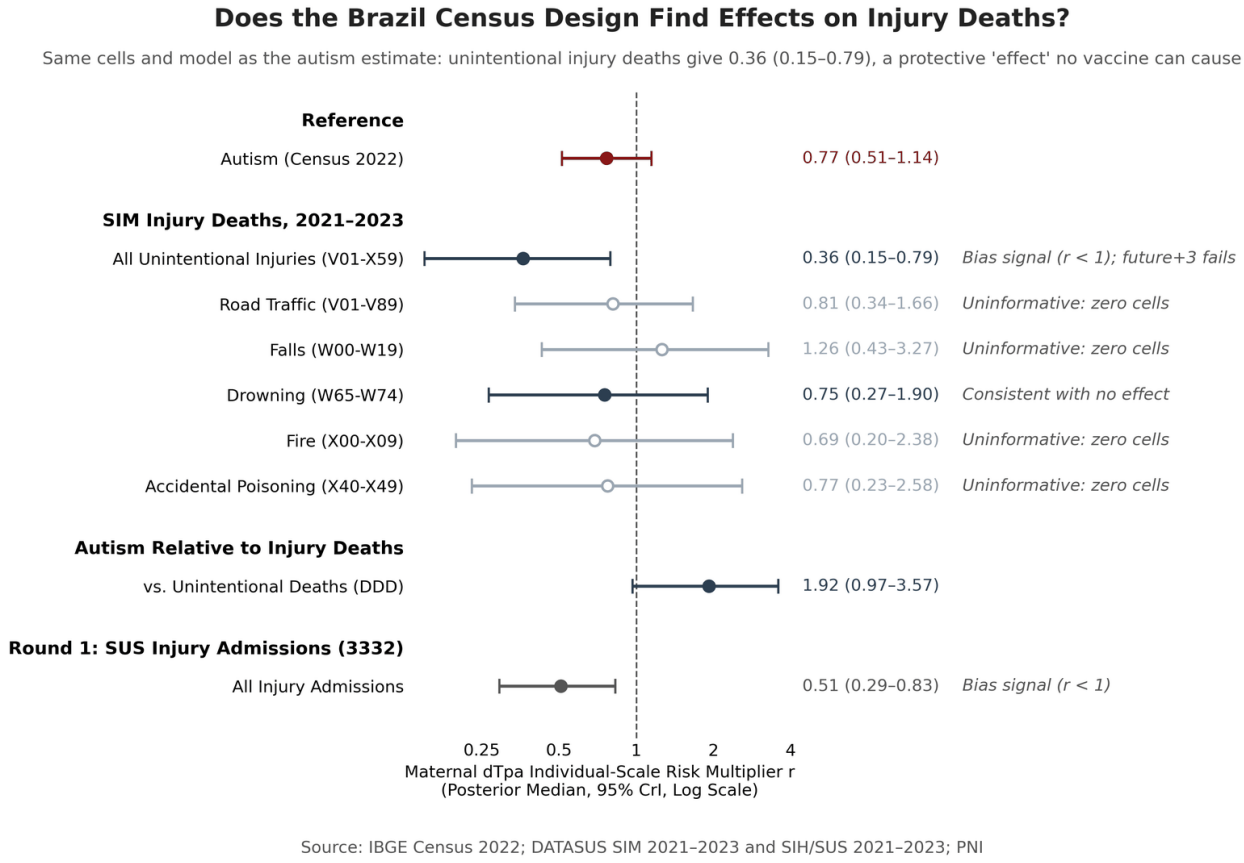


Figure S98. Brazilian Census maternal DTaP design with SIM injury deaths. Autism, unintentional injury deaths and their components in death years 2021-2023 by age at death, the autism-relative DDD and the round-one SUS admission composite on the same 108 cells. Hollow light markers are UNINFORMATIVE (zero-death cells). Unintentional injury deaths are a protective-direction bias signal with a failing future+3 placebo, and the reading is DESIGN BIAS DETECTED (r < 1).

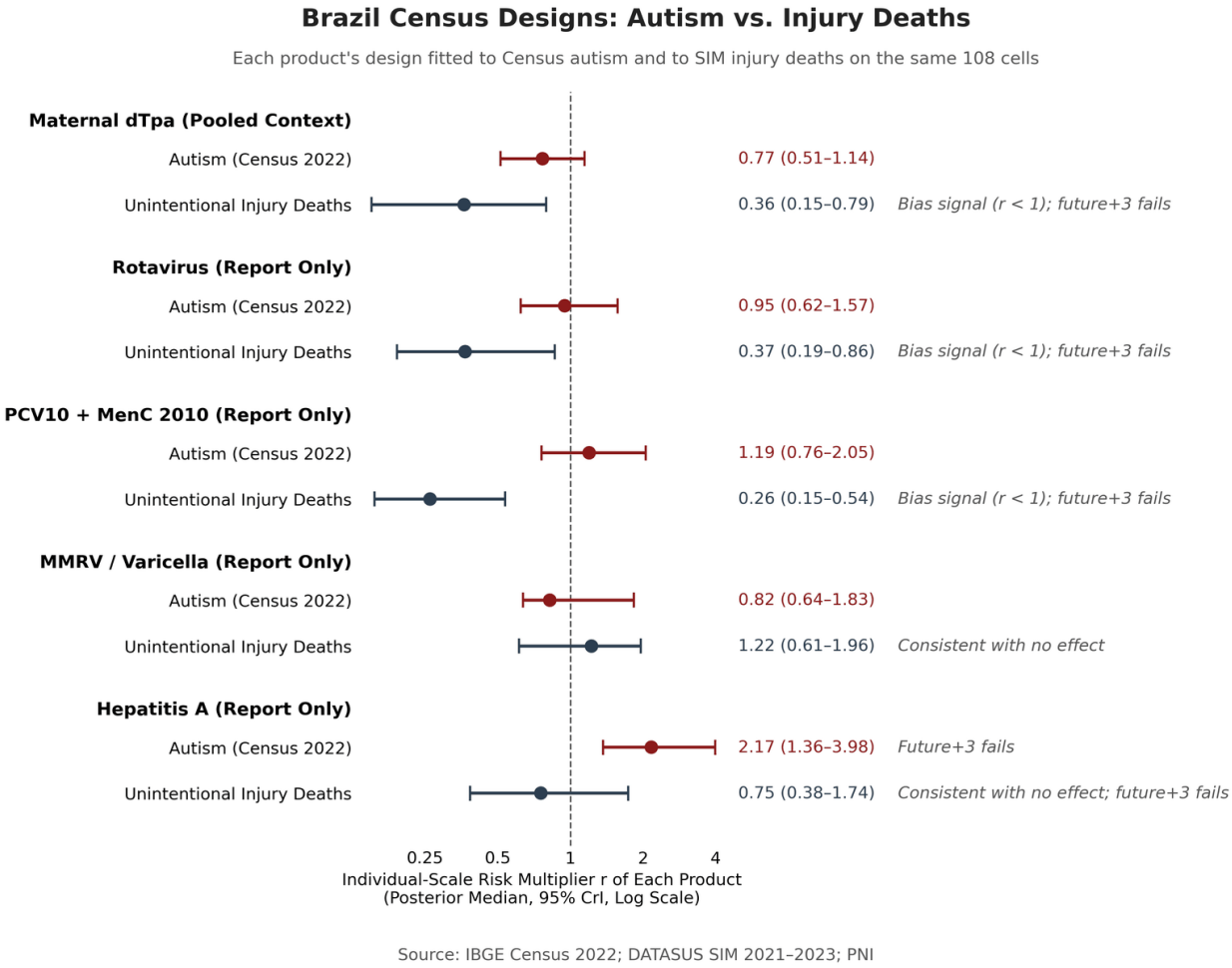


Figure S99. Brazilian Census designs by product with SIM injury deaths. Autism and unintentional injury deaths under the maternal DTaP, rotavirus, PCV10 with MenC, MMRV and hepatitis A designs on the same 108 cells. Injury deaths are protective-direction bias signals under the maternal DTaP, rotavirus and PCV10 with MenC designs. The five designs share cells and counts and are not independent tests.

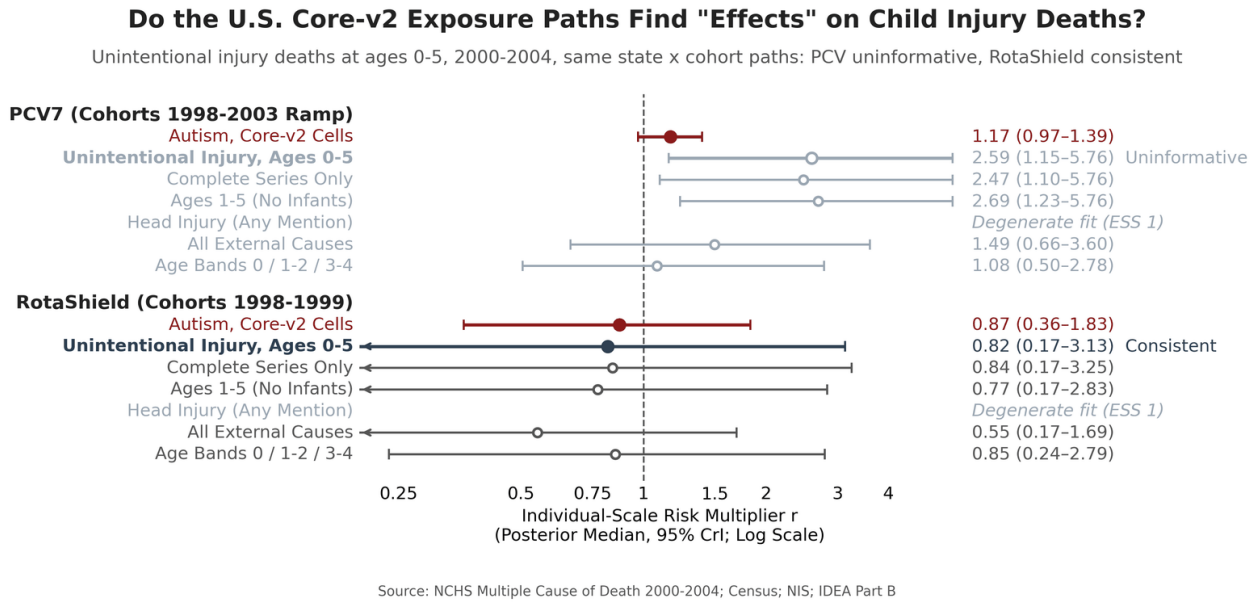


Figure S100. U.S. core-v2 exposure paths with NCHS child injury deaths. Unintentional injury deaths at ages 0-5 in 2000-2004 on state-by-birth-cohort Lexis cells, with secondary cell rules and outcomes, beside the core-v2 autism posteriors (red) for PCV7 and RotaShield. Light hollow markers are UNINFORMATIVE; dark hollow markers are secondary fits classed CONSISTENT. The PCV control is UNINFORMATIVE by the width rule although its interval excludes one, and its future+3 placebo fails; RotaShield is CONSISTENT but prior-dominated. The head-injury fits are degenerate (effective sample size 1).

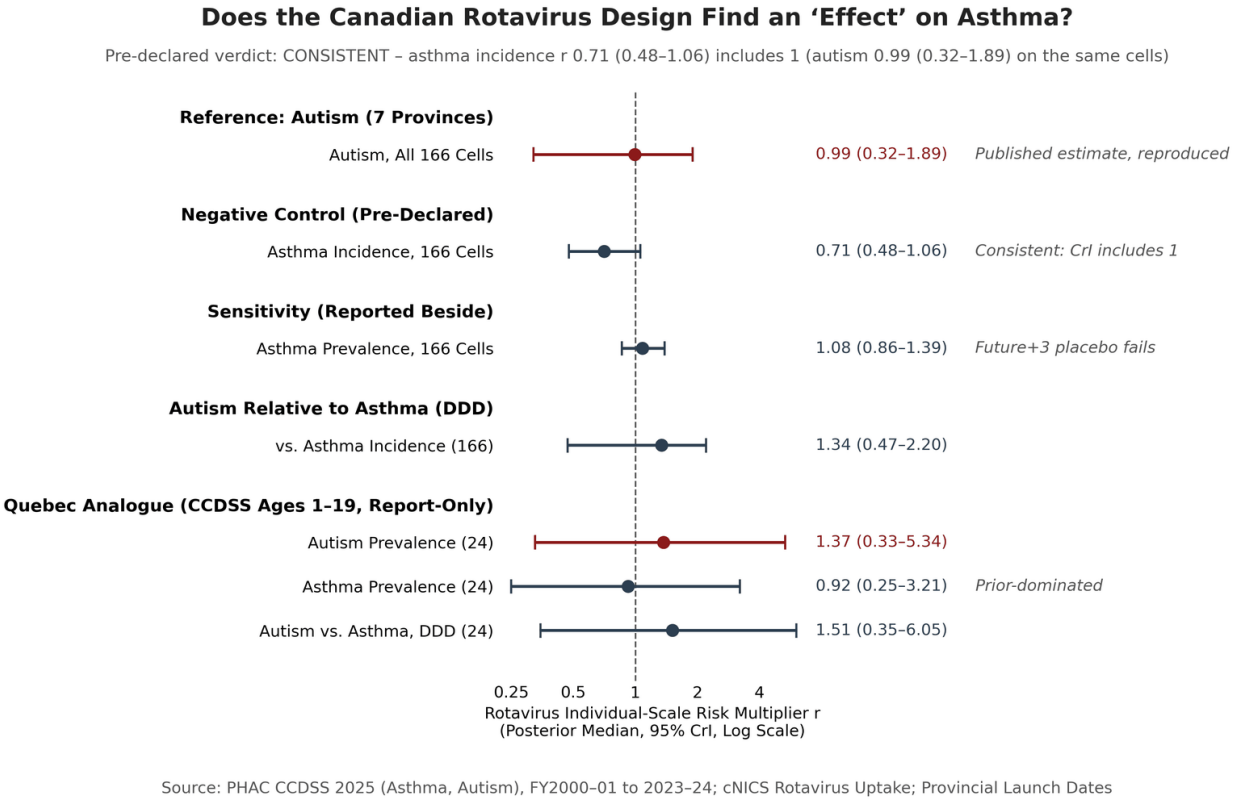


Figure S101. Canadian rotavirus design with CCDSS asthma. Autism on all 166 cells of the design excluding Quebec, the stated asthma-incidence control, asthma prevalence (a sensitivity), the autism-relative DDD and a report-only Quebec analogue on 24 CCDSS cells. Asthma incidence is CONSISTENT with an interval narrower than autism's; the prevalence sensitivity fails its future+3 placebo.

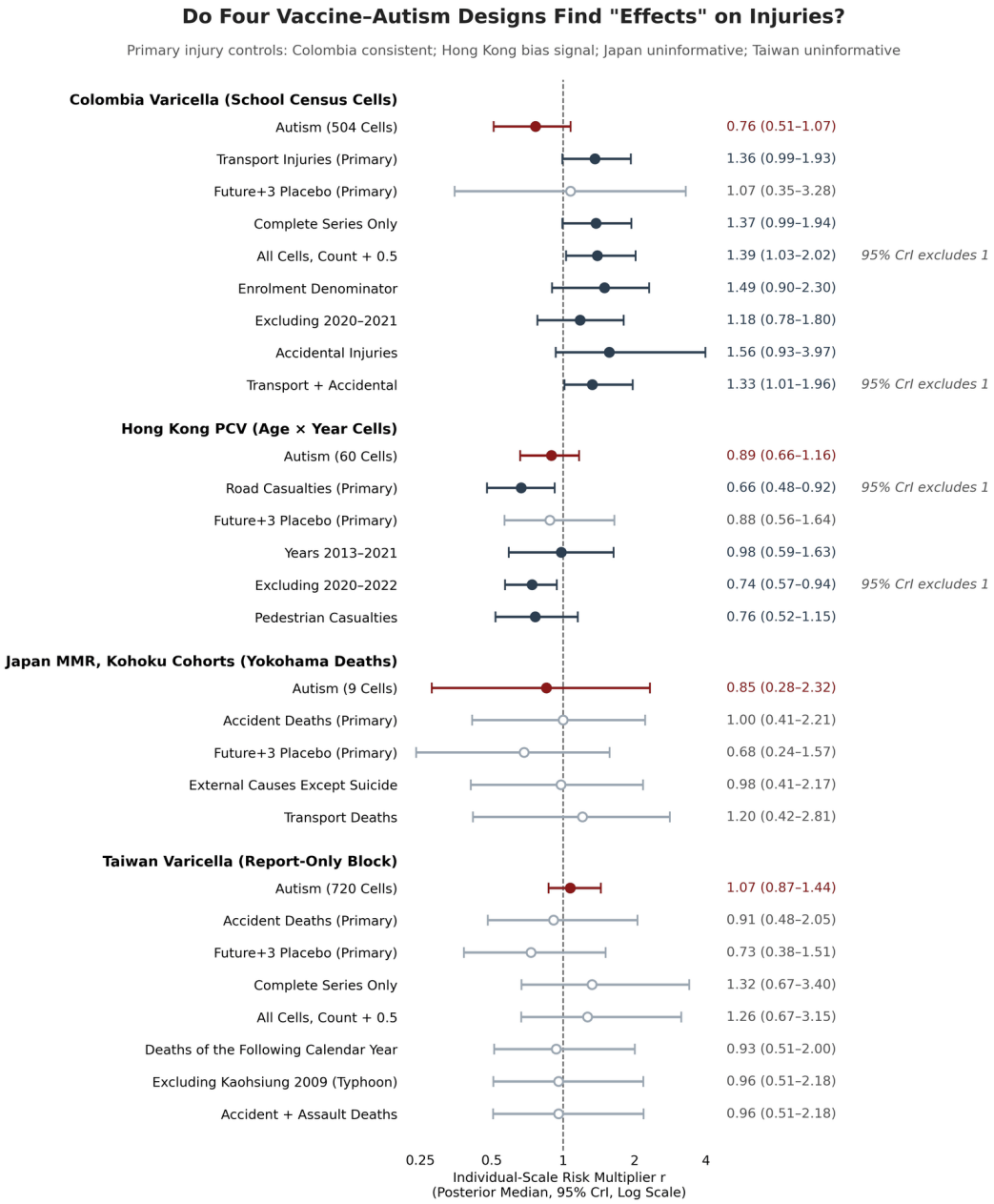


Figure S102. Colombian, Hong Kong, Japanese and Taiwanese designs with round-two injury controls. Autism on its own cells (red) beside INMLCF transport and accidental injuries (Colombia), road-traffic and pedestrian casualties (Hong Kong), Yokohama deaths at ages 15-19 (Japan MMR) and MOHW accident deaths (Taiwan), with future+3 placebos and sensitivities. Hollow light markers are UNINFORMATIVE controls and placebo rows. The Colombian primary is CONSISTENT with a lower limit of 0.991, Hong Kong road casualties are a protective-direction bias signal, and the Japanese and Taiwanese controls are UNINFORMATIVE.

Files: `3326_us_core_negative_control_outcomes.md`, `.csv` and `.json`; `3331_injury_negative_controls_us.md` through `3348_injury_negative_controls_germany_destatis.md` with their `.csv` and `.json` outputs; round-two files `3380_injury_negative_controls_uk_round2.md`, `3381_injury_negative_controls_europe_round2.md`, `3382_injury_negative_controls_brazil_sim.md`, `3383_injury_negative_controls_us_deaths.md`, `3384_negative_controls_canada.md` and `3385_injury_negative_controls_col_hkg_jpn_twn_round2.md` with their `.csv` and `.json` outputs; the synthesis `3330_negative_control_summary.md` and `.csv`, regenerated with round two (`qa/negative_controls_20260925/3330_ROUND2_UPDATE.md`); rule files in `qa/negative_controls_20260925/` and `qa/stage2b_20260925/`.

S64. Interactions, schedule load, meta-regressions and convergence

S64.1 The jump-interaction layer

For a pair of vaccines, theta is the autism-rate multiplier for a child who received both, beyond the product of the two single-vaccine rate ratios; theta = 1 is log additivity and theta > 1 is harmful stacking. The layer pools log theta over exchangeable pairs in a normal-normal model with mean prior N(0, (log 2)^2) and between-pair SD prior half-normal(0.5), evaluated exactly on a 6,001-point grid. Its mean, theta_int, is reported beside the shared mean and is never multiplied into it. The rule `qa/layers_20260929/3360_rule.json` was written before any combined posterior was computed.

The inputs are 27 pairs from three sources. The 24 U.S. state pairs are the eligible core rows of the cross-state interaction file, disattenuated for errors in NIS-measured joint receipt (pair-specific attenuation where simulated, otherwise 0.71) and correlated through state influence functions (off-diagonal correlations -0.62 to 0.80; 51 state clusters). Two English sequences, rotavirus followed by MenB and MMR followed by rotavirus, enter with calibrated standard errors and the checker's jackknife correlation of -0.2606. The U.S. core PCV-by-RotaShield term enters as its likelihood after the core prior is divided out, N(-1.452, 1.648^2); the joint U.S. likelihood barely informs it. The layer reproduces the U.S. derived pool (0.9719629868, 0.8836258707-1.070598487) and the English independence pool (0.8499441531, 0.4983499798-1.457308183) to machine precision.

Table S64-A. Interaction hyper-mean theta_int under the primary layer and every stated sensitivity.

Scenario	Pairs	theta_int (95% CrI)	P(theta_int > 1.10)	P(theta_int > 1.25)
Primary	27	0.967 (0.881-1.063)	0.0046	2.7e-06
U.S. state pairs only	24	0.972 (0.884-1.071)	0.0070	7.1e-06
England pairs only	2	0.871 (0.446-1.733)	0.2381	1.4e-01
U.S. core term only	1	0.815 (0.231-2.873)	0.3203	2.5e-01
Without the influenza-by-Hib-deferral pair	26	0.952 (0.868-1.042)	0.0014	8.6e-07
Without U.S. state pairs	3	0.831 (0.429-1.602)	0.1899	1.0e-01
Without England pairs	25	0.971 (0.883-1.069)	0.0065	6.1e-06
Without the U.S. core term	26	0.969 (0.882-1.064)	0.0049	3.1e-06
England pairs independent	27	0.966 (0.880-1.061)	0.0041	2.3e-06
England pairs maximally dependent	27	0.968 (0.881-1.065)	0.0050	3.3e-06
Adding England MMR (five years) followed by PCV	28	0.968 (0.882-1.064)	0.0048	2.8e-06
No attenuation correction (lambda 1)	27	0.975 (0.912-1.044)	0.0006	3.4e-08
Stronger attenuation (lambda 0.62)	27	0.965 (0.870-1.069)	0.0070	7.9e-06
U.S. pairs independent	27	1.019 (0.906-1.138)	0.0876	2.4e-04
Adding hybrid-tier U.S. rows	32	0.974 (0.886-1.070)	0.0064	3.3e-06
Between-pair SD prior half-normal(log 1.25)	27	0.967 (0.882-1.062)	0.0042	1.8e-06
Between-pair SD prior half-normal(1)	27	0.967 (0.881-1.063)	0.0046	2.9e-06
Mean prior scale halved	27	0.968 (0.882-1.063)	0.0044	2.3e-06
Mean prior scale doubled	27	0.967 (0.880-1.063)	0.0046	2.8e-06
U.S.-alone core term	27	0.967 (0.881-1.063)	0.0045	2.7e-06

Notes. Posterior medians with 95 percent credible intervals, exact on the grid. No scenario shows a synergy signal, defined by the rule as a 95 percent interval excluding one in the harmful direction. Source: `3360_interaction_layer.csv`, row_type pooled.

The primary theta_int is 0.967 (0.881-1.063), with P(theta_int > 1) = 0.239, P(> 1.10) = 0.0046 and P(> 1.25) = 2.69e-06. Between-pair heterogeneity is small (tau 0.094, 0.025-0.192), and a new exchangeable pair is predicted at 0.967 (0.762-1.229), with probability 0.020 of exceeding 1.25. At the tau median the U.S. state pairs carry 97.5 percent of the information, England 2.4 percent and the U.S. core term 0.08 percent. Under the estimated U.S. correlation four rows carry small negative GLS weights, together -0.083; treating the U.S. pairs as independent is the largest move among sensitivities that keep every source, to 1.019 (0.906-1.138). Without the U.S. state pairs only three weak inputs remain, 0.831 (0.429-1.602), consistent with the primary but unable to rule out moderate synergy on their own.

The one pair whose interval excludes one is influenza by Hib booster deferral, 1.373 (1.085-1.737) as checked and 1.50 (1.11-2.03) after errors-in-variables correction (p 0.009). It does not survive multiplicity across the 38 primary U.S. rows (Benjamini-Hochberg q = 0.354). Its excess sits in a Hib-booster main effect of 0.76 (0.63-0.92) among children without influenza vaccine; among influenza recipients the booster effect times theta is 1.04. It coincides with the 2009 H1N1 season and the December 2007-June 2009 Hib shortage, and its time-shift placebos do not reproduce it. The layer neither hides this pair nor lets it decide the result.

Several items are listed but not pooled. The Japanese pairs are not identified because routine coverage saturates in every prefecture and joint receipt is unobserved (for example Japanese encephalitis at age three by Hib/PCV, 1.037 (0.786-1.368)). The Swedish and German pairs are not identified (permutation SDs of at least 1.07, imputation-dependent hepatitis B exposure, or estimability only under a Frechet-lower rule with placebos equal to the estimate). The ADDM varicella-by-PCV7 term, 0.654 (0.223-1.822), is a separate panel overlapping the state population. The Hib booster deferral is a timing estimand, reported in Section S60.4. The 55 excluded U.S. state rows, including RotaShield by PCV7, are not identified, not separable or superseded.

The layer is pairwise and area-level. Theta is identified from state or area variation in joint coverage converted through the mixture equation; joint receipt is measured in survey samples or bounded and is never linked to the same child's outcome. The layer therefore does not establish same-child biology, does not test three-way or higher-order stacking, dose spacing or total load, and covers pairs never studied here only through exchangeability. The older ecological schedule-interaction hierarchy, 1.009 (0.953-1.074), measures a population-scale departure from log additivity per simultaneous ten-point coverage shift in both vaccines (Section S39.3). It reuses the same U.S. and English outcome series and answers a different question, so the two are never combined.

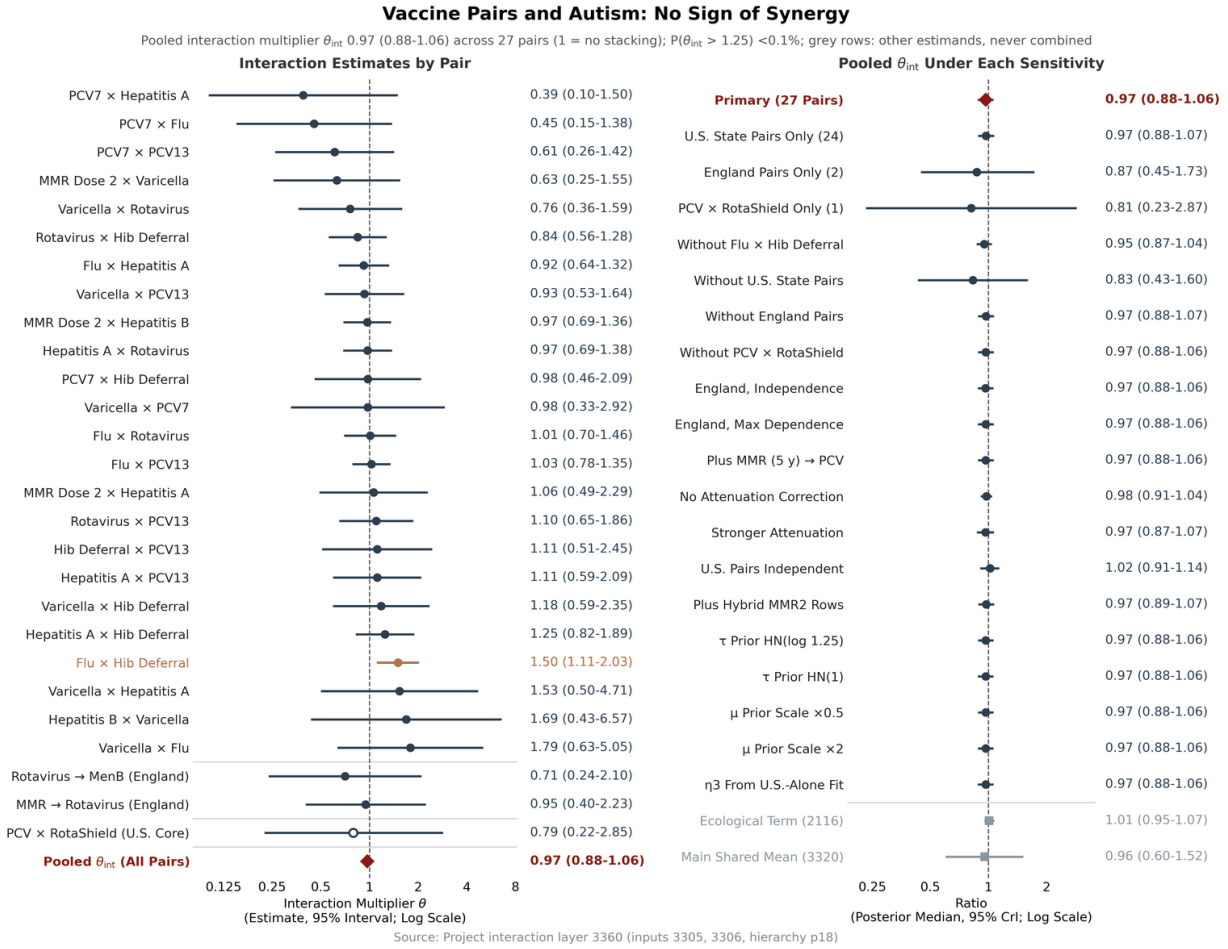


Figure S103. Jump-interaction layer. Left: interaction multipliers for the 24 U.S. state pairs, two English sequences and the U.S. core PCV-by-RotaShield term, with the pooled hyper-mean; the influenza-by-Hib-deferral pair is highlighted. Right: the pooled hyper-mean under every stated sensitivity, beside the older ecological interaction term and the main shared mean, which are different estimands and are never combined. A multiplier of 1 means no stacking.

S64.2 Schedule load and age at exposure

The schedule-load file is an exposure dataset, not an effect estimate: 583 schedule-change rows across 17 jurisdictions after checker corrections, and a birth-year panel of 701 rows on 18 tracks giving the load in the first 24 months as antigens, injections, doses and micrograms of aluminium. It has no likelihood, theta, individual-scale r, placebo or load-dependence test, and none was created for it.

The meta-regressions supply the load test. Each pooled hierarchy context with a load in that file enters once through its own likelihood curve, with the hierarchy's product and context random intercepts, its bias SDs (0.35 or 0.70) and its shared group terms. The load is the number of distinct antigens, or of injections, scheduled at or before the product's first dose in the regime the product joined. The slope prior is N(0, (log 2 / 10)^2) per antigen or injection, so one prior SD is a doubling or halving over ten units. Ten contexts qualify, spanning 5 to 6.29 antigens and 1 to 5 injections; Hong Kong, Czechia, Brazil, Colombia and Tanzania have no verified load and are left out rather than imputed. The rule `qa/layers_20260929/3361_rule.json` was written before any statistic was computed.

Table S64-B. Load and age meta-regressions.

Variant (contexts)	Metric	Ratio (95% CrI)	Prior 95% range	Information ratio
A1 primary (10)	per +5 antigens	0.999 (0.507-1.965)	0.507-1.972	0.996
A1 primary (10)	per +5 injections	0.982 (0.516-1.865)	0.507-1.972	0.945
A2 adding U.S. rotavirus 2006 (11)	per +5 injections	0.965 (0.544-1.712)	0.507-1.972	0.843
A3 0-24 month load (10)	per +5 injections	1.018 (0.620-1.669)	0.507-1.972	0.728
A4 no bias terms (10)	per +5 injections	1.101 (0.712-1.648)	0.507-1.972	0.614
A5 doubled slope prior (10)	per +5 injections	0.945 (0.308-2.895)	0.257-3.891	0.823
A7 adding two U.S. PCV7 survey jump rows (12)	per +5 injections	0.996 (0.530-1.872)	0.507-1.972	0.928
C1 age at first dose (10)	per +10 months	0.906 (0.321-2.523)	0.322-3.102	0.910

Notes. The information ratio is the posterior SD divided by the prior SD; the stated rule classes a slope as NOT INFORMATIVE when it is at least 0.8. Antigen rows for A2-A7 and the age sensitivities C2 and C3 are in the source table; every interval includes one. Savage-Dickey Bayes factors for a zero slope, descriptive and not stated, lie between 1.00 and 1.48. Source: `3361_meta_regressions.csv`.

Per five additional antigens the ratio is 0.999 (0.507-1.965), with P(slope > 0) = 0.499, and per five injections 0.982 (0.516-1.865). The posterior barely moves from the prior (information ratios 0.996 and 0.945), and the verdict for both is NOT SUPPORTED; NOT INFORMATIVE. The variants in which the data move the prior most, removing every bias term and using the 0-24 month load, still include one. The ratio per ten months of age at first dose is 0.906 (0.321-2.523), also not informative; only Japan MMR (12 months) and Sweden PCV (3 months) differ from the two-month products, so the age slope rests mainly on the weak Japan likelihood and is confounded with product. With ten contexts, four products and loads clustered at 5-6.3 antigens and 1-5 injections, a null slope is absence of evidence, not evidence that load has no effect. Load is a property of the schedule, not of the individual child.

S64.3 Same-product consistency

Contexts of the same product agree on their own data, without bias allowances. Modern rotavirus (five contexts) has tau 0.09 (0.00-0.50) and Cochran Q = 1.93 (df 4, p = 0.749); adding U.S. rotavirus 2006 gives Q = 1.94 (df 5, p = 0.857). PCV (four contexts) has tau 0.16 (0.01-0.67) and Q = 2.64 (df 3, p = 0.450), and MMR (two contexts) tau 0.30 (0.01-1.03) and Q = 1.03 (df 1, p = 0.310). No context conflicts with the rest of its product in leave-one-out cross-validation; the smallest conflict p is 0.108. U.S. PCV, 1.17 (0.97-1.39) on its own data, against the other PCV contexts pooled at 0.95 (0.79-1.14), has conflict p = 0.108 under homogeneity, 0.442 with random effects and 0.828 with the assumed bias SDs, and is CONSISTENT by the stated rule. In a two-level model of the fifteen pooled contexts the within-product SD is 0.069 (0.000-0.240) and the between-product SD 0.102 (0.000-0.408). Neither spread is distinguishable from zero, but two to six contexts per product cannot pin down heterogeneity, so agreement means no detectable conflict, not proven homogeneity.

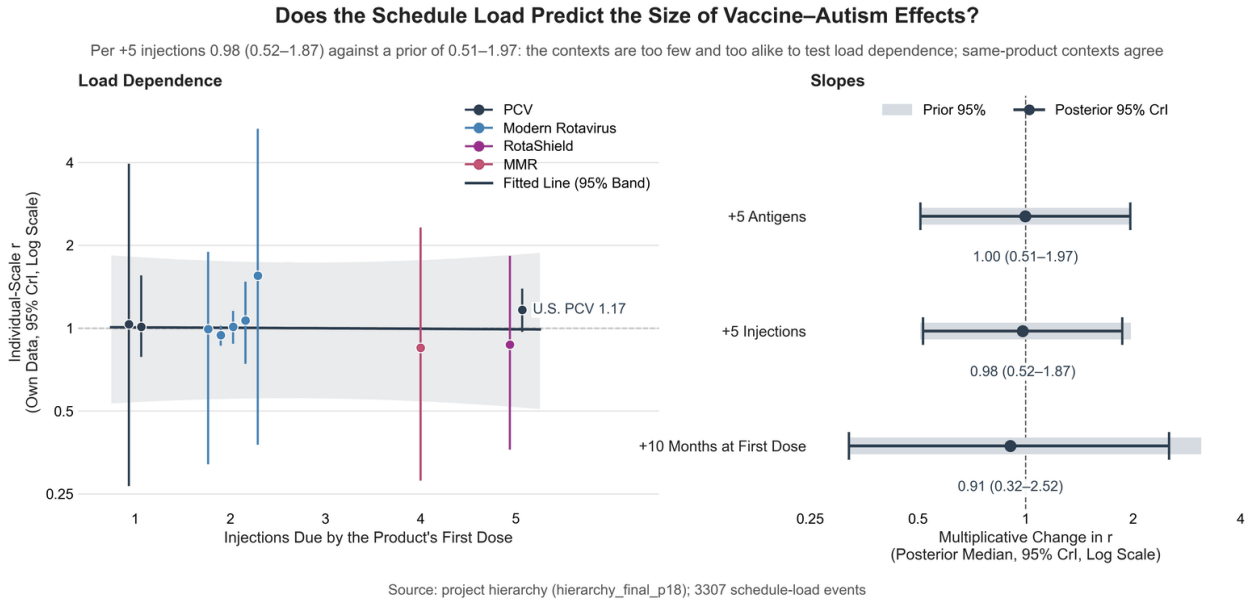


Figure S104. Schedule load and age at exposure. Left: each pooled context's own-data r against the injections due by the product's first dose, with the fitted load slope and its 95 percent band. Right: posterior slopes per five antigens, per five injections and per ten months of age at first dose against their prior 95 percent ranges. The posteriors barely move from the priors.

S64.4 Convergence across designs

The convergence panel (Figure S107) places, for each drawn context, the hierarchy's own-data posterior beside every jump, synthetic control, native-scale design, individual study from another population and negative control on the individual rate-ratio scale. Its universe is 601 rows in 55 contexts. Twenty contexts are drawn, the fifteen pooled contexts plus U.S. rotavirus 2006, the Hib booster deferral, U.S. hepatitis A 1999, Japan's Japanese encephalitis block and Sweden rotavirus, with 108 drawn rows. Main-text Figure 4 shows nine of these contexts: the same-cell comparisons of England, Northern Ireland and Canada rotavirus, Sweden PCV, U.S. hepatitis A 1999 and U.S. rotavirus 2006, and the three contexts whose designs show a bias signal on a primary negative control, U.S. PCV, Brazil maternal DTaP and Hong Kong PCV. Its negative-control layer adds files 3347, 3348 and 3380-3385, read through the harmonized rows of file 3330 (Section S63.13), to the round-one task files: the 16 round-two reading controls on identical or closest cells are drawn, and the rows of files 3347 and 3348 (national injury deaths and German 2024 discharges) are listed but not drawn. Rows without a placebo-passing specification, judged uninformative or artefactual by the checker, report-only, exploratory, not eligible, or classed UNINFORMATIVE as negative controls are drawn hollow and light; dashed whiskers mark the 90 percent intervals of the Canadian synthetic controls. Of the drawn design rows, 23 use the same cells as the context's own data, 9 the same children in different cells or series and 5 overlapping cells. Agreement among them is robustness to estimator and functional form, not a second draw of data, and it cannot rule out a bias they share. No row in the panel enters the shared mean.

Where designs share cells they agree. England rotavirus gives 0.94 (0.86-1.02) on its own data and 0.95 (0.82-1.10) as a jump, drawn hollow because no specification passes its placebo check. Northern Ireland gives 1.01 (0.88-1.16) and 0.91 (0.73-1.15) on the same 155 cells. Sweden PCV gives 1.01 (0.79-1.56) and 0.98 (0.77-1.26) in the all-ages F84 block. U.S. hepatitis A 1999 leans above one in the jump, 1.12 (0.69-1.82), the stacked synthetic difference-in-differences, 1.18 (0.61-3.24), and the dose-response version, 1.23 (0.81-1.78), each interval including one. Japan MMR gives 0.85 (0.28-2.32) on its own data beside the MMR cohort meta-analysis, 0.96 (0.88-1.05), and its injury-death control, 0.86 (0.46-1.61).

Where designs disagree the disagreement is diagnostic. The drawn Canadian synthetic controls sit at 1.76-4.99 with every lower bound at zero, against 0.99 (0.32-1.89) from the hierarchy on the same cells, because only ages one to three are exposed in the clean window (Section S61.3); Quebec gives 1.07 (0.74-1.48) against a synthetic control of 1.81 (0.00-7.05). U.S. rotavirus 2006 gives 0.95 (0.76-1.19) untied in the core and 1.31 (0.74-2.33) as a jump on the same IDEA cells, leaning in opposite directions with overlapping intervals, and its negative controls carry leans of their own (traumatic brain injury 2.37, 1.09-4.58; hearing impairment 1.41, 1.07-1.82). In the U.S. core, Brazil and Hong Kong

panels the same designs return "effects" on candidate control outcomes (in Hong Kong in both directions; Section S63), so same-cell agreement there cannot establish specificity.

The panel also lists how far the shared-mean upper limit moves when one context is left out: 1.511 without U.S. PCV (-0.006), 1.527 without Canada (+0.010), 1.539 without Hong Kong (+0.022), 1.547 without England (+0.030), 1.560 without Brazil (+0.043) and 1.633 without Colombia (+0.116), against the primary's 1.517. Substituting Colombia's lockdown-free curve instead (file 3371; Section S62.8) moves it to 1.655 (+0.138).

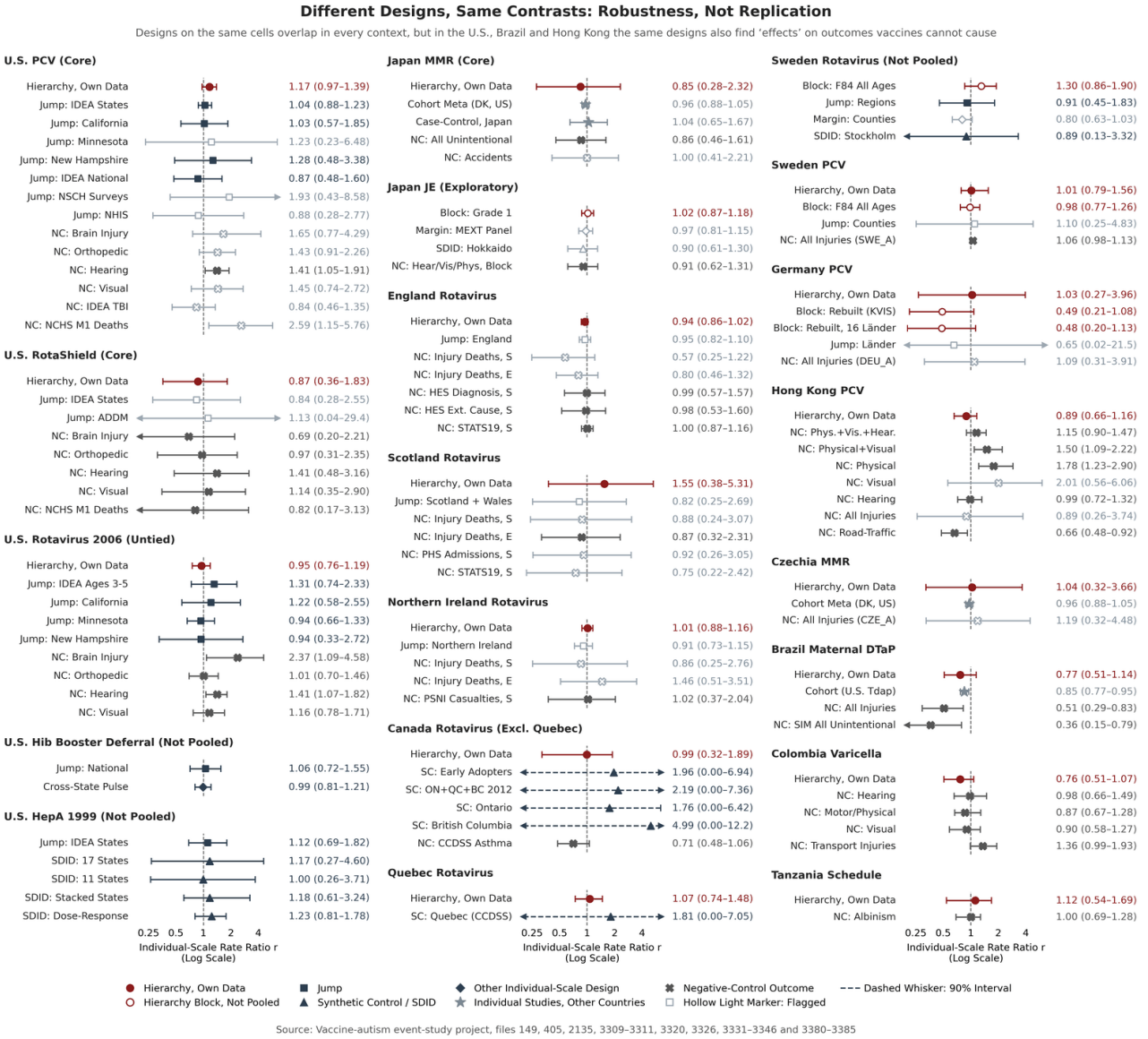


Figure S107. Different designs on the same vaccine-by-place contrasts, all 20 drawn contexts (file 3362). Each panel shows, on the individual rate-ratio scale r (log axis), the context's own-data hierarchy posterior (filled red circle), a not-pooled hierarchy block of the same series (hollow red circle), jump estimates (squares), synthetic controls and synthetic difference-in-differences (triangles), other individual-scale designs (diamonds), individual-level studies of the same product in other populations (stars) and negative-control outcomes (crosses), with 95 percent intervals; arrows mark intervals that extend beyond the axis. Hollow light markers are flagged as lacking a placebo-passing specification, uninformative, report-only or exploratory; dashed whiskers are 90 percent intervals. The 108 rows cover the 15 pooled contexts, U.S. rotavirus 2006 (untied), the Hib booster deferral, U.S. hepatitis A 1999, the Japanese encephalitis block and Sweden rotavirus. Designs on shared cells overlap in every context, but in the U.S., Brazil and Hong Kong panels the same designs also return "effects" on outcomes vaccination cannot cause, so agreement is robustness, not replication. No row enters the shared mean. Main-text Figure 4 shows nine of these contexts.

Files: `3360_interaction_layer.md`, `.csv` and `.json`; `3305_interactions_us.md`; `3306_interactions_uk_japan_other.md`; `3307_schedule_load_results.md`; `3361_meta_regressions.md`, `.csv` and `3361_meta_regressions_inputs.csv`; `3362_convergence_panel.md` and `.csv`.

S65. Control calibration: residual bias, dependence and held-out contexts

This post hoc sensitivity uses the full fifteen-context calibrated likelihood and adds no independent observations. It retains the specified-bias primary synthesis. Negative-control calibration requires both an unaffected outcome and a defensible relation between its bias and autism's bias; outcome labels do not establish either condition (Lipsitch, Tchetgen Tchetgen and Cohen, 2010; Schuemie et al., 2018). Asthma, sensory and broad physical disability are also claim-relevant outcomes, not guaranteed biological nulls. Injury controls can still be affected indirectly. An apparent control effect therefore challenges specificity under the exclusion restriction rather than proving that an autism association is confounding.

S65.1 Residual autism-specific bias

Let a = theta + b, where theta is the common log risk ratio and b is additional shared autism-specific bias, independent of the control model, with distribution N(0, s squared). We integrate the calibrated likelihood L(a) against that distribution and the original theta prior N(0, (log 2) squared). Values of s are sensitivity assumptions, not estimates from the control data. V1 retains the control-derived mean shifts; V2 keeps only their variances.

Table S65-A. Calibrated common-effect sensitivity to additional shared autism-specific bias.

Calibration	Additional log-bias SD	Common RR (95% CrI)	P(RR > 1.25)
V1, mean shifts retained	0.00	0.871 (0.741-1.020)	<0.001
V1, mean shifts retained	0.15	0.876 (0.632-1.212)	0.016
V1, mean shifts retained	0.30	0.890 (0.510-1.551)	0.115
V2, variance only	0.00	0.946 (0.806-1.111)	<0.001
V2, variance only	0.15	0.949 (0.685-1.314)	0.049
V2, variance only	0.30	0.955 (0.547-1.665)	0.171

Source: `3392_calibration_stress_tests.json`. The full grid also includes SDs 0.05, 0.10 and 0.20. Exclusion of a common 25 percent increase is conditional on the calibration and residual-bias assumptions. None of these summaries bounds every product or subgroup. Medians below one do not establish a protective effect. These computations inherit the low local importance-sampling effective sizes of file 3370; agreement between its two scrambles is a diagnostic, not a numerical error bound, particularly for small tail probabilities.

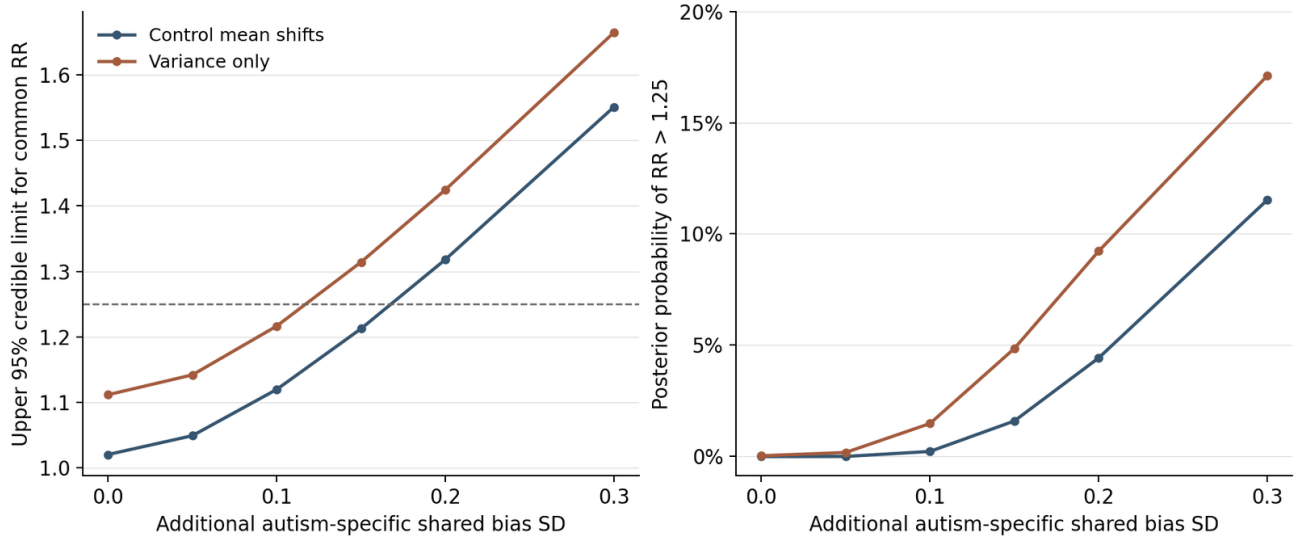


Figure S110. Additional autism-specific shared bias after control calibration. Left: upper 95 percent credible limits for the common individual-scale RR; the dashed line marks 1.25. Right: posterior probability of a common RR above 1.25. V1 uses control-derived mean shifts and variances; V2 uses variances only. The bias SD is assumed, not estimated. These post hoc scenarios use the same fifteen contexts and inherit the calibrated fits' numerical limitations (S65).

S65.2 Dependence among controls

The calibration-input model is refitted under covariance C = diag((1 - rho) s_j squared + omega squared) + rho s s transpose within each context, for assumed correlations rho = 0, 0.5 and 0.9. The covariance concerns sampling error; outcome-specific bias deviations remain independent. Four sets are examined: all 45 primary controls, the existing rule-literal de-nested set of 38, its 25 injury controls and its 13 non-injury controls. The latter subsets contain five and six contexts, respectively. Correlations are sensitivity values, not identified by the marginal standard errors.

At rho = 0.5 the posterior mean bias is +0.035 (posterior SD 0.073) in the de-nested set and +0.071 (SD 0.121) for its injury controls. For the injury set at rho = 0.9 these become +0.117 and 0.178. These are control-error-model diagnostics, not corrected autism estimates: the autism hierarchy has not been refitted under these covariance assumptions. Nor do the within-context scenarios resolve dependence between the two U.S. exposure paths. Positive correlation need not monotonically widen the hierarchical result because the estimated bias components also change.

S65.3 Whole-context deletion and verification

We remove every control from a context before calibrating its held-out estimates, removing both U.S. exposure paths together. At rho = 0.5, null coverage is 0.974 across the de-nested rows and 0.984 when held-out groups receive equal weight (eight groups). It is 1.000 by both summaries for the injury subset (five groups). These descriptive coverage rates use dependent outcomes and few groups; they are not independent-binomial validation rates and do not select a preferred model. Broad intervals can cover the null without establishing bias exchangeability with autism.

The zero-residual-bias summaries reproduce the original V1 and V2 common-effect outputs within an absolute tolerance of 1e-9. The rho-zero error likelihood reproduces the original diagonal model. Unit tests compare the covariance implementation with a direct multivariate-normal likelihood and verify that correlated duplicate controls do not provide independent precision. The analysis records source and cache hashes. Files: `3392_calibration_stress_tests.py`, `test_3392_calibration_stress_tests.py`, `3392_calibration_stress_tests.json`, and `3392_calibration_stress_tests.md`.

S66. Joint vaccine contrasts within shared outcome records

This secondary, post hoc extension fits multiple exposure changes to the original outcome records jointly. Brazil contributes 108 state-by-age census cells and three contrasts: rotavirus, the 2010 PCV10/MenC schedule composite, and maternal DTaP. The hepatitis A and problematic MMRV series remain nuisance columns. Sweden contributes PCV, hepatitis B and rotavirus exposures to either the 756 existing age 5-14 cells or 1,122 age 0-14 cells. The larger age set replaces, rather than supplements, the smaller one. This common-center comparison represents 19 contrasts in the same 15 contexts and 14 likelihood blocks, with 5,499 original cells. These are not 19 independent experiments.

S66.1 Target and joint receipt assumptions

Under independent receipt conditional on an outcome cell, the log-rate shift is the sum across products of log(1 - p_j + p_j exp(theta_j)). Each cell enters one normal-inverse-gamma likelihood, using the same six trend and serial-dependence kernels as the international analysis. All exposure scenarios are averaged at the likelihood level: seven synchronized Brazil scenarios and 24 combinations of Sweden's hepatitis B and rotavirus scenarios. Brazil's 2010 coverage average is a composite proxy, not an observed probability of joint receipt. Maternal vaccination is a maternal-exposure contrast, not an infant dose.

The within-country log effect is theta_j = a + b_country + d_j. Independent product deviations d_j have assumed normal SD 0, 0.35 or 0.70; they can represent differential effect or differential bias, which these data do not separate. Integrating them out gives a joint country marginal likelihood for a + b_country. A single country bias then enters under the existing assumed allowances. The resulting country factor replaces the original maternal-DTaP or Swedish-PCV factor before global shared-bias integration. Neither country's other vaccine factors were previously in the synthesis. The core likelihood, Germany/Sweden care-system bridge and other contexts are retained. This common-center comparison is distinct from the 17-context heterogeneous hierarchy in S67, which also uses joint country likelihoods and adds survey-outcome blocks.

Marginal uptake does not identify which children received combinations. An equal-effect sensitivity therefore uses maximally positively dependent receipt, assigning all products by one shared uniform rank in each cell. This model and

the independence model do not exhaust possible receipt patterns. Neither estimates a biological interaction parameter. The target is a common center per represented vaccine or schedule contrast, not a universal individual risk ratio, a per-dose effect, or a bound on all disabilities. Country and product bias decomposition remains assumption-dependent; additional product-specific dependence between countries is not identified here.

S66.2 Synthesis and product-specific results

Table S66-A. Joint-country common-center sensitivity.

Model	Product-deviation SD	Shared-bias SD	Common center (95% CrI)	P(>1.25)
Existing common effect	0.00	0.15	0.964 (0.637-1.455)	0.108
Joint, expanded Sweden	0.00	0.15	0.979 (0.647-1.476)	0.122
Joint, expanded Sweden	0.35	0.15	0.980 (0.648-1.480)	0.124
Joint, expanded Sweden	0.70	0.15	0.980 (0.647-1.481)	0.124
Joint, nested receipt	0.00	0.15	0.979 (0.648-1.477)	0.122
Joint, expanded Sweden	0.35	0.30	0.983 (0.539-1.789)	0.216

The 0.35-deviation row's log-interval width differs from the existing common-effect model's by less than 0.01 percent. More contrasts improve the completeness of the likelihood, but do not materially improve precision. The full files contain all 42 synthesis scenarios, including single-country substitutions and the smaller Swedish age set. Eight prior sensitivities fix product-deviation SD at 0.35: neutral log-scale SD log(2), concentrated SD log(1.25), wide SD log(4), and a harm-centered prior with mean log(1.2) and SD log(1.5), each under shared-bias SD 0.15 and 0.30. None excludes a common 25 percent increase at its 95 percent upper limit. These priors are sensitivity assumptions, not extra data.

The uncalibrated joint product marginals are not all centered on one. In Brazil, rotavirus gives 0.899 (0.560-1.555), the 2010 schedule 1.294 (0.777-2.486), and maternal DTaP 0.762 (0.504-1.149). In expanded Sweden, PCV gives 0.994 (0.771-1.290), hepatitis B 1.297 (0.940-1.738), and rotavirus 1.292 (0.844-1.906). These diagnostics use independent log-effect priors with SD log(2) and do not include the synthesis-level country bias allowance. They retain both positive and negative associations; none establishes causal harm or protection.

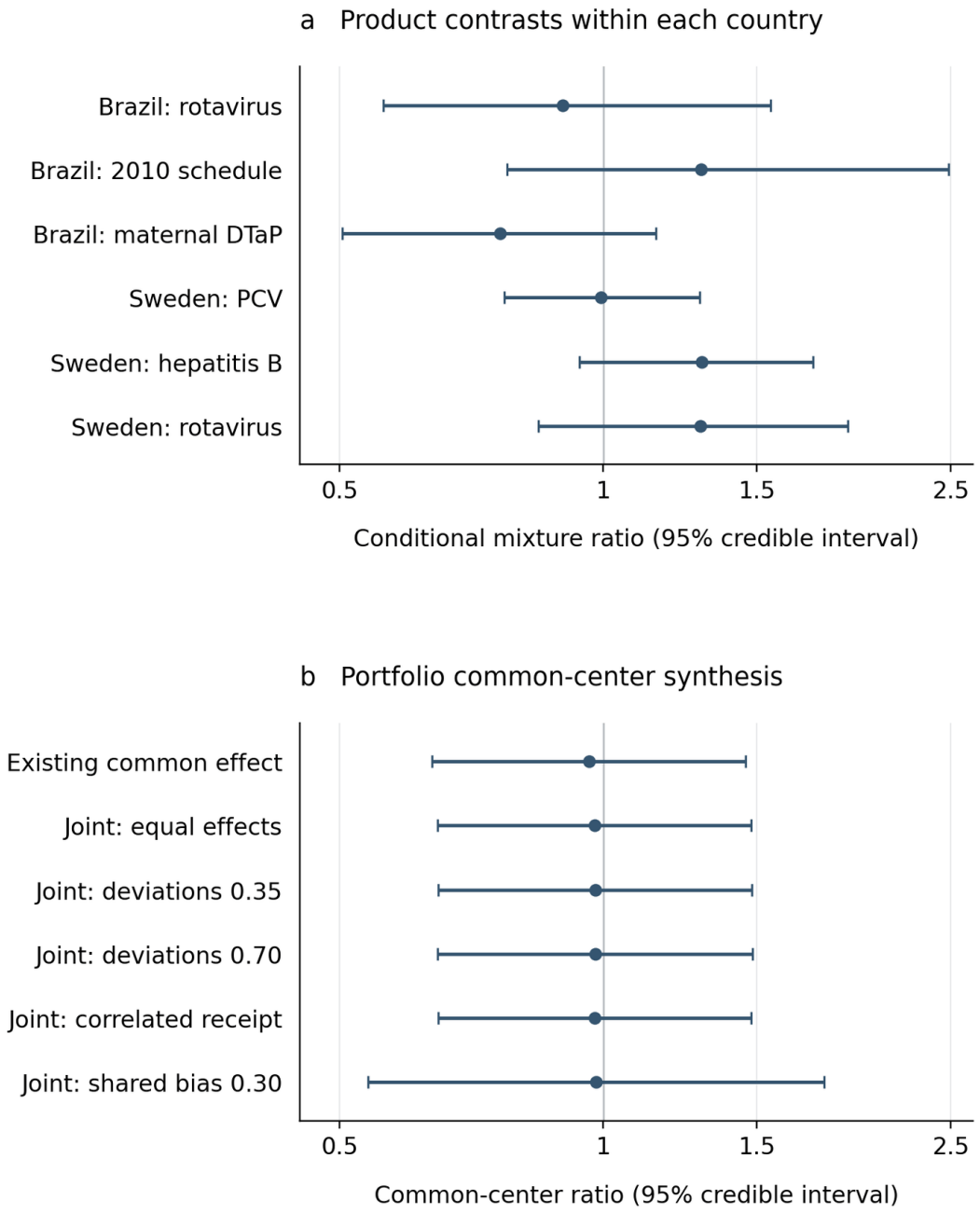


Figure S111. Joint vaccine contrasts without duplicated outcome records. Panel a shows joint product marginals in Brazil and expanded-age Sweden, before synthesis-level country bias allowances. Panel b compares the existing common effect with secondary joint-country replacements. Product deviations and shared bias are log-scale SDs. All rows except the last use shared-bias SD 0.15; the last uses product-deviation SD 0.35 and shared-bias SD 0.30. Correlated receipt is the equal-effect nested model. Rows reuse observations; they are not independent studies or biological interaction estimates (S66).

S66.3 Diagnostics and numerical verification

All twelve future-three-year placebo summaries include one. These assignments can remain correlated with actual exposure, so null inclusion does not establish identification. Noise-free injections at ratios 1, 1.25, 1.5 and 2 yield monotonically increasing fitted medians under every kernel summary. Brazil recovers the endpoints as 1.066 and 2.203, indicating imperfect absolute calibration. Sweden's endpoints are 1.007 and 2.012 for the existing age set, and 1.002 and 2.002 for the expanded set. These are recovery diagnostics under simulated no-bias conditions, not empirical power estimates.

The three-dimensional product grid uses 161 points from -6 to 6. Comparing 121 points on the same range changes the synthesis median, endpoints and P(>1.25) by at most 0.0000015. Extending the range to -7.5 to 7.5 with 201 points at the same spacing changes them by less than 0.000000000001. Both pass the specified 0.005 tolerance. Product-deviation convolution uses zero padding and an eight-SD Gaussian kernel without renormalizing truncated mass. The saved original common-effect result is reproduced, and an independent frozen-likelihood replay matches all 42 summaries to the stated 1e-9 tolerance. Six unit tests verify alignment, direct versus quadratic likelihoods, receipt special cases, analytic convolution, baseline reproduction and full factor-substitution algebra. Local core effective sizes and two-scramble differences are retained in the machine-readable table; numerical agreement does not establish causal identification.

Files: `3376_joint_country_synthesis.md`, `.csv`, `.json`; `3376_joint_country_synthesis.py`; `test_3376_joint_country_synthesis.py`; `processed/report_joint_country_3376.py`; and `project_results/joint_country_3376/`. No published-cohort meta-analysis, new independent observational source, or control-calibrated hierarchy refit is claimed by this extension.

S67. Whole-project integration on compatible scales

S67.1 Evidence units and estimands

The integrated report at `project_results/whole_project_synthesis_3377/index.html` links the original-likelihood autism model, interaction and cumulative-exposure models, local native-scale families, published-study families and condition-specific service-use models. These are complementary targets, not independent estimates of one universal vaccine risk. The frozen register contains 60,844 overlapping result records, 531 registry entries and 519 original result files. All have dispositions; this establishes accounting for the acquired register, not worldwide data exhaustion or identification of every recorded association. The extended register adds source-verified contrasts and separately labeled derived results. Its machine-readable receipt gives the release-specific count.

The main autism hierarchy represents 21 vaccine or schedule contrasts across 17 contexts in 16 likelihood blocks, with 8,538 original outcome cells. Brazil and Sweden enter through joint multiple-product likelihoods, replacing their single-product contributions rather than adding copies of their outcome records. NSCH and NHIS contribute separate survey-outcome likelihoods. U.S. PCV and RotaShield retain their joint original-observation likelihood; U.S. rotavirus 2006 remains estimated but untied from the shared mean. Different measurement systems are not presumed to be independent causal replications: shared U.S. bias and shared-country sensitivities explicitly examine dependence beyond the product hierarchy.

The target is the exponentiated shared mean log effect across represented contrasts. Its interpretation requires the exposure model, counterfactual trends, outcome ascertainment and assumed bias distribution. It is not the effect of receiving the complete schedule rather than no vaccines, nor a population-weighted average over all children. Products whose timing is collinear contribute a schedule contrast, not separately identified product effects. No number of secondary specifications adds independent likelihood information when it reuses the same observations.

S67.2 Autism synthesis and Bayesian sensitivity

Table S67-A. Original-likelihood synthesis and specified sensitivities.

Model	Mean ratio	95% credible interval	P(mean > 1.25)
Joint products plus NSCH and NHIS, reference	0.987	0.638-1.519	0.139
Same model, shared-bias SD 0	0.987	0.699-1.381	0.080
Same model, shared-bias SD 0.30	0.988	0.536-1.819	0.225
Same model, mean-prior SD log 4	0.986	0.625-1.545	0.148
Joint products without survey blocks	0.991	0.638-1.532	—
Restricted 15-context comparison	0.957	0.599-1.517	0.126

The reference continuous prior for the shared log mean is normal with SD log 2, the global shared-bias SD is 0.15, and the shared U.S. bias SD is 0.35. Removing the global shared term does not remove context-specific bias allowances. The joint country and U.S. dependence variants, heterogeneity settings and source-specific exposure scenarios are

retained in the linked tables. The intervals permit moderate increases under the reference assumptions. Similar medians across specifications cannot establish that systematic error is absent.

The Bayes factor for a zero shared mean against the reference continuous alternative is 3.181. Prior point-null masses of 0.10, 0.20, 0.50, 0.80 and 0.90 yield posterior point-null probabilities of 0.261, 0.443, 0.761, 0.927 and 0.966, respectively. These values test whether the average log effect is zero while retaining product and context heterogeneity under both hypotheses. They are not probabilities that all effects are zero. The continuous posterior gives P(mean > 1.10) = 0.310 and P(mean > 1.25) = 0.139. Negative posterior mass likewise does not prove prevention; a near-zero mean can mask offsetting effects or shared bias.

The separate interaction model gives 0.967 (0.881-1.063) across 27 contrasts in three working overlap blocks. It concerns an average departure from log additivity for the represented exposure pairs. The cumulative-load contrast is 0.999 (0.507-1.965) per five additional antigens and remains largely prior-driven. Neither result supplies a total-schedule risk bound. The restricted 15-context common-effect, control-calibration and power calculations in S62-S66 use their stated likelihoods and are not operating-characteristic validations of the expanded hierarchy. In particular, the calibrated 0.871 (0.741-1.020) requires bias exchangeability between autism and controls; it is not the headline estimate in Table S67-A.

Sources: `autism_synthesis_comparison.csv`, `joint_dependence_sensitivity.csv`, `shared_mean_probability_sensitivity.csv` and `shared_mean_point_null_sensitivity.csv` in the integrated report directory; `07_analysis_code/3377_joint_heterogeneous_synthesis.py`.

S67.2a Likelihood support without the shared-mean prior

A likelihood-ratio diagnostic removes the top-level prior on the shared mean, averages the two numerical integrations in likelihood space, and profiles over the six saved product/context heterogeneity choices. It retains the internally marginalized nuisance distributions, exposure bridges and working likelihood. Shared additive Gaussian bias is integrated at each stated SD before profiling; those SDs are assumptions, not empirical bias estimates.

Table S67-D. Conditional likelihood-ratio support sets for the exponentiated shared mean.

Shared-bias SD, log scale	Maximizing ratio on saved grid	Nominal 95% support set
0	1.000	0.728-1.342
0.15	1.000	0.647-1.511
0.30	1.000	0.510-1.919

The chi-square cutoff has not been calibrated for frequentist coverage in this integrated model. These are therefore nominal support sets, not validated 95% confidence intervals, individual risk ratios or simultaneous bounds on every vaccine effect. Fifty-four repeated integration and nuisance specifications form a numerical sensitivity set, not additional observations. They also show why removing the shared-mean prior does not justify a narrow universal exclusion bound. Source: `qa/data_handling_20261002/shared_mean_likelihood_support_3576/run_001/likelihood_support_sets.csv`, using the 241-point survey grid, numerical-average and profiled-heterogeneity rows; `07_analysis_code/shared_mean_likelihood_support_3576.py`.

S67.3 Local coverage gradients by denominator and exposure

The local analysis audits 22 designs from 21 states. Seventeen use autism's share of special education and five use enrollment denominators. Their exposure measures include kindergarten MMR, toddler MMR, school protection proxies and complete-schedule compliance. The historical nine-design calculation combined five kindergarten-MMR special-education composition equations, three enrollment-based outcomes, and Wisconsin's complete-schedule compliance. Its 0.992 (0.980-1.004) interval therefore cannot be interpreted as a common MMR population-risk estimate. Wisconsin's dominant weight describes that heterogeneous calculation, not the relative information of nine comparable MMR replications.

For autism counts A, special-education counts S and population or enrollment N, the identity is log(A/N) = log(A/S) + log(S/N). An effect on autism's share of special education does not determine the effect on population prevalence unless the corresponding effect on S/N and its covariance are available. Holding the missing term at zero is a substantive assumption, not an algebraic conversion. The all-child individual-risk inversion of the mixed nine-design coefficient has no valid interpretation here.

Seven exploratory families each receive seven summaries: a modified Hartung-Knapp-Sidik-Jonkman calculation, a sampling-only Bayesian fit, a reference bias fit and four bias/dependence sensitivities. Only same-year coefficients enter these syntheses. Their log-mean prior has SD log 1.25 and heterogeneity prior is half-normal with scale 0.10. Reference design-specific and shared-bias SDs are both 0.025 per ten coverage points. Sensitivities use shared-bias SD 0.05, sampling correlations 0.5 and 0.99, or both design-specific and shared-bias SD 0.10 with a wider log-mean prior SD log 2. These choices are specified working assumptions, not estimated confounding or covariance. Working standard errors do not narrow the original small-cluster intervals: each uses the larger of the reported log standard error and the normal-equivalent interval width.

Table S67-B. Local native-scale reference summaries per ten coverage percentage points, conditional on specified working priors rather than estimated bias allowances.

Exposure and outcome denominator	States	Ratio	95% credible interval
Kindergarten MMR; autism share of special education	12	0.999	0.946-1.054
Same target, original screened subset	5	0.994	0.922-1.073
Same target, omitting failed diagnostic	11	0.997	0.943-1.053
Same target, federal California source substituted	12	1.001	0.948-1.057
Toddler MMR; autism share of special education	2	1.019	0.890-1.162
Schedule compliance; autism share of special education	2	0.984	0.885-1.096
School protection proxy; enrollment-based outcome	3	1.085	0.925-1.354

These families overlap and are never combined as independent summaries. California's state and federal outcomes are alternatives, not two states. Massachusetts kindergarten MMR with an enrollment denominator remains standalone, ratio 1.012 (original 95% CI 0.969-1.057); New Jersey schedule compliance with an enrollment denominator is also standalone, 1.012 (0.884-1.160), with changing numerator age scope. Oregon's baseline-relative conversion holds the observed baseline fixed. Suppression, age mismatch and endogenous uptake remain limitations. Failed timing diagnostics and diagnostics unavailable in short panels are distinguished; neither a missing test nor a nonrejecting test certifies identification.

For the twelve-state kindergarten-MMR composition target, raising shared-bias SD to 0.05 gives 0.999 (0.905-1.102). Raising both study-specific and shared-bias SDs to 0.10 while widening the mean prior gives 0.998 (0.813-1.224). Assumed sampling correlations 0.5 and 0.99 give 1.000 (0.944-1.058) and 1.003 (0.950-1.058). These comparisons quantify how the narrow reference interval depends on the bias assumptions; none validates those assumptions or supplies an individual-risk exclusion bound.

Files: `local/local_native_effects.csv`, `local/local_native_synthesis.csv` and `local/summary.json` in the integrated report directory; `07_analysis_code/3396_local_native_synthesis.py`.

S67.4 Outcome-year alignment and geographic checks

Illinois and Vermont lagged coverage analyses join school enrollment for the outcome year before shifting exposure. Exposure-year enrollment weights the county coverage measure but does not replace the outcome-year count-model offset. Illinois county aggregation uses 102 county identifiers rather than 113 combined regional-office/county codes. These changes preserve the outcome definition and estimator; they repair alignment, not identification.

The source-linked refit set contains 1,708 Illinois and 720 Vermont main-effect rows and twenty interaction rows in each state. Across 610 district same-year comparisons, coefficients and intervals reproduce the frozen calculations to numerical precision. Twelve frozen release hashes remain unchanged. For the primary Illinois school-protection contrast, future exposure gives 0.979 (0.900-1.064), p = 0.616; contemporaneous, one-year and two-year ratios are 1.018, 0.993 and 1.023. Vermont gives future exposure 1.201 (0.353-4.085), p = 0.767; its contemporaneous ratio remains positive, 3.431 (1.846-6.376), while one- and two-year ratios are 1.462 and 1.509 with wide intervals. All five Vermont interaction future-placebo tests reject; none of the five Illinois tests rejects. The positive Vermont result is retained, not converted to a null because other models fail.

Broad-school-age outcomes, top-coded coverage, suppression exclusions, diagnosis-policy confounding and incomplete cells still preclude an individual causal interpretation. The complete refits, comparisons and tests are linked through `qa/integration_closure_20261002/local_lag_repairs.json` and `local_lag_refits/`. They are alternative specifications on existing observations, not new independent populations.

S67.5 Published studies and overlapping outcome contrasts

The published-study census contains 4,683 contrast records with 85 specified participant/source identifiers. Thirty-seven family definitions use 48 distinct contrasts, with repeated membership handled as alternative fits rather than independent likelihood contributions across families. HR, RR, OR and IRR remain distinct; where a bridge is considered, it is labeled as a sensitivity. Observational autism estimates are not inserted into the original population-policy likelihoods. Self-controlled acute safety estimates answer narrow timing questions, not chronic autism or total-disability questions. These identifiers include composite multi-database sources and do not imply 85 independent samples.

The Jain et al. article corrected on 12 January 2016 supplies ten adjusted age/dose/familial-risk contrasts in Table 2. Its methods and footnote identify the reported adjusted ratios as Cox hazard-rate ratios, despite the article's RR label. Values include one MMR dose at age two with an affected older sibling, HR 0.76 (0.48-1.22), and two doses at age five with an unaffected older sibling, 1.09 (0.76-1.54). They come from one interaction model with shared participants and unknown cross-contrast covariance. Each of ten alternative MMR-HR fits adds exactly one of these rows to Hviid and Burstain; their means range 0.903-0.977 and every interval includes one. These are mixed-target sensitivities: exchangeability between age/dose/familial-risk conditional hazards and the broader cohort targets is assumed, not established. Ten contrasts do not become ten cohorts. Extraction uses the corrected JAMA article copy hosted by the Minnesota legislature, with the publisher's correction notice linked; the file hash establishes byte integrity, not publisher authentication.

Jain's Optum index births are 2001-2007. Burstain's Epic Cosmos cohort enters at ages one to two during 2015-2024, so the index-child birth windows do not overlap. Different claims and EHR systems alone would not have established that. Their shared U.S. selection and ascertainment mechanisms remain represented by the same-country bias term and covariance sensitivity. Jain's reference state is unvaccinated at the relevant age, with time-varying dose status; it is not a group containing earlier-vaccinated children. Burstain's reported interval is 99 percent, and that level is used when recovering its log standard error.

Table S67-C. Selected published-study reference summaries.

Target and design	Inputs	Estimate	95% credible interval
Childhood MMR and autism, cohort HR	2	0.954	0.647-1.415
Childhood MMR, IRR/HR bridge sensitivity	3	0.947	0.694-1.293
Prenatal influenza and autism, cohort HR	4	1.012	0.779-1.315
Childhood MMR, case-control OR	2	0.926	0.585-1.513
PCV13, febrile seizure days 0-1 versus 14-20, IRR	3	1.770	1.116-2.720
Inactivated influenza, same risk-window IRR	3	1.350	0.845-2.193
IIV plus PCV13 coadministration, risk-window IRR	2	3.027	1.413-5.403
COVID mRNA, 21-28-day seizure bridge, Japan dose 1	3	1.077	0.696-1.645
Same bridge, Japan dose 2 substituted	3	1.081	0.700-1.658

The reference native-family model uses log-mean prior SD log 2, half-normal heterogeneity scale 0.35, study-bias SD 0.10, same-country system-bias SD 0.05 and global shared-bias SD 0.10. Sampling-only, larger bias, assumed same-country sampling correlation, wider priors and leave-one-study-out variants are retained. A family has at most one selected contrast from each specified participant spine. No product of overlapping family posteriors is reported. The IIV-plus-PCV13 estimate compares post-vaccination risk and control windows, not same-day with separated administration; it is not a biological synergy estimate. The COVID bridge deliberately varies window length, age mix, formulation and dose across countries and is exploratory. Japan's two dose rows are substitutes, not independent additions.

Lee et al. (2025) supplies 114 additional preperiod, psychiatric-history, age, sex and platform contrasts from Tables 2-4, alongside its six overall post-vaccination rows. All belong to the same Korean source study. Positive history-free subgroup estimates include anxiety/nervousness IRR 1.17 (1.15-1.18) and sleep disorders 1.07 (1.06-1.09), despite inverse overall estimates. Pre-vaccination associations are diagnostics, not vaccine effects; subgroup significance does not remove time-varying confounding or multiple-testing concerns. The printed age-19-or-younger sleep interval, 0.98 (0.87-0.10), is internally inconsistent and remains flagged and unfitted. No replacement endpoint is invented. These rows make the retained evidence more complete without being counted as independent replications or evidence of universal psychiatric safety.

The publisher's Supplementary Tables 2-11 add 288 contrast records, including dose, individual vaccine, exclusion and seven-day-window analyses. All thirteen tables are preserved as 2,231 structured cells, including outcome definitions and diagnostic counts. Among 84 repeated main-table targets, 79 numerical triples agree and five disagree: second-income-quartile anxiety, and cognitive impairment by residential location and vaccine platform. Both versions remain unpooled and source-linked; the supplementary document repeats the malformed age-19-or-younger sleep interval and cannot repair it. Timing findings include sleep disorders at days 1-7, IRR 1.06 (1.05-1.08), and anxiety at days 15-21, 1.03 (1.02-1.05). These are exploratory, overlapping short-window associations, not established chronic harms, and they are not suppressed because the overall estimates are inverse. Files `literature/lee_supplement_contrasts.csv`, `lee_supplement_all_table_cells.csv` and `lee_main_supplement_crosscheck.csv` preserve the extraction and discrepancies; `07_analysis_code/lee_supplement_3399.py` and its tests reproduce them.

Primary sources: Jain et al. (2015), corrected Table 2, DOI `10.1001/jama.2015.3077`, with correction DOI `10.1001/jama.2015.17065`; Lee et al. (2025), Tables 2-4 and Supplementary Tables 1-13, DOI `10.1192/bjo.2025.10803`. Files: `literature/all_native_study_dispositions.csv`, `literature/native_family_synthesis.csv`; `07_analysis_code/published_study_updates_3395.py` and `07_analysis_code/3379_published_native_synthesis.py`. Local source PDFs, the supplementary DOCX and hashes are included in the reproduction manifest.

S67.6 Condition-specific outcomes and reproducibility

The cross-country service-use layer keeps speech/language disorders, motor developmental disorders, ADHD and epilepsy separate. Their reference mean ratios are 1.116 (0.437-2.817), 0.909 (0.329-2.519), 0.983 (0.412-2.347) and 1.004 (0.416-2.417), respectively. These broad intervals and institutional outcome definitions do not establish absence of disability effects. Acute adverse-event evidence is also not a negative control for autism: it concerns other phenotypes and latency windows and must be interpreted on its own merits.

`07_analysis_code/3394_replay_whole_synthesis.ps1` provides verification and model-replay entry points. The integrated report's `reproducibility_manifest.json`, `completion_checks.csv` and `completion_receipt.json` record input hashes, tested identities and execution scope. Tests check source-value updates, same-spine exclusions, denominator alignment, original-model reproduction, profile and country-grid refinement, and independent integration scrambles. Passing these checks establishes specified computational identities, not the validity of untestable causal assumptions or a raw-source replay of every historical file. Positive, negative, uninformative, incompatible and technically blocked sources retain explicit dispositions.

S68. Broader inclusion and acute-safety bridges

S68.1 International autism panels

The broader-inclusion model adds five geographic contexts and replaces one overlapping German panel. It uses the archived original-observation likelihood profiles, not their posterior medians or multiple specifications as independent evidence. The resulting 22 contexts represent 26 vaccine or schedule contrasts in 21 likelihood blocks and 11,641 reporting cells. The primary remains the 17-context model in S67.

Table S68-A. Additional or replacement autism panels and diagnostic limitations.

Context and exposure	Cells	Role	Failed diagnostics
Wales, rotavirus	220	Additional context	Pretrend
Poland, PCV	416	Additional context	Pretrend
South Korea, schedule	918	Additional context	Pretrend, placebo, harm injection and null simulation
Taiwan, varicella	720	Additional context	Placebo
Chile, PCV	40	Additional context	Null simulation
Germany, PCV	825	Replaces 36-cell German series	Null simulation

Education classification, claims and hospital encounters measure different recorded outcomes. Chilean and German discharge counts are episodes, not unique autistic people. Wales has a short near-unexposed preperiod and a change in special-education reporting. Germany reconstructs early coverage and harmonizes claims and later surveillance coverage. These limitations remain in the archived source notes; a model label does not remove them.

Each added or replacement log-effect profile receives an independent zero-centered normal panel-bias allowance with SD 0.70 or 1.40. Wales also shares the existing UK-system bias; Germany replaces the original German contribution within the German/Swedish specialist-care term. The added PCV, varicella and schedule contexts attach to their existing product terms. The original shared U.S. bias, product/context heterogeneity mixture, joint country likelihoods and global shared-bias term remain. Germany is never an additional independent population alongside its smaller series.

These allowances are sensitivity assumptions, not estimates of bias. Pretrend failure may produce directional, dependent or non-Gaussian errors that these priors do not cover. Including such panels cannot turn failed identification into validated causal evidence. The two broader-inclusion estimates reuse the primary observations and are not independent meta-analyses to combine again.

Table S68-B. Broader-inclusion shared-mean sensitivity.

Model	Ratio	95% credible interval
Primary, 17 contexts	0.987	0.638-1.519
22 contexts, added-panel bias SD 0.70	0.932	0.612-1.413
22 contexts, added-panel bias SD 1.40	0.961	0.624-1.470

The table uses shared-mean prior SD log 2 and global shared-bias SD 0.15. Twelve broader-inclusion specifications cross the two panel-bias SDs, mean-prior SDs log 2/log 4, and global bias SDs 0/0.15/0.30. Medians range 0.926-0.965, but the most permissive interval extends to 1.852. Thus the additional panels do not establish equivalence or exclude moderate effects.

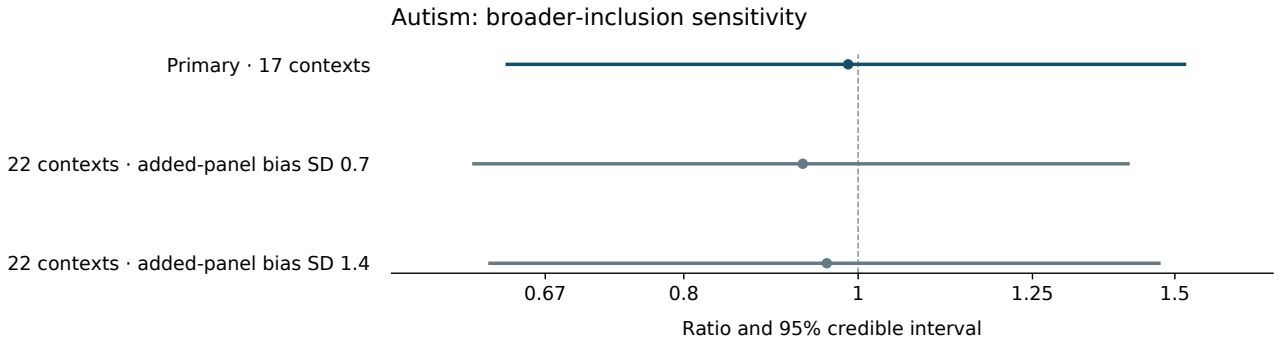


Figure S112. Broader international autism inclusion. The 22-context sensitivities add five contexts and replace the smaller German series. Added panels failed pretrend, placebo or simulation diagnostics. Intervals are conditional on the stated normal bias allowances and retain substantial uncertainty. Rows overlap and are not independent evidence (S68).

Two power-18 randomized quasi-Monte Carlo integrations differ by less than 0.003 in interval endpoints across these specifications. Relevant importance-sample effective sizes exceed 118. Refining the country integration grid from 161 to 201 points changes the six SD-0.70 summaries by less than 0.000002. A neutral-factor identity test reproduces the original model's message factorization to numerical precision. These are computation checks, not causal-validity tests.

Files: `07_analysis_code/3400_expanded_inclusion_synthesis.py`, `test_3400_expanded_inclusion.py`, and `expanded_inclusion/` under the integrated report directory. Source dispositions retain the diagnostic failures and input hashes. The portable replay recomputes all twelve saved-likelihood probability scenarios, while the full working-copy replay can rebuild their integrated likelihoods from held country and survey profiles.

S68.2 Published short-window febrile seizures

Two explicitly bridged acute-safety families use three previously unpooled contrasts. They do not enter the autism likelihood or establish chronic disability risks.

Table S68-C. Acute-safety bridge sensitivities under the native-family reference priors.

Exposure and bridge	Inputs	Ratio	95% credible interval
MMR, Australian and Chinese dose/window bridge	2	1.459	0.807-2.436
PCV, US and Korean formulation/window bridge	4	1.528	1.062-2.266

Macartney et al. (2015) report Australian first-dose MMR at 12 months, with days 5-12 relative incidence 1.9 (1.3-2.9). Xu et al. (2021) report Ningbo MMR at 18 months after measles-rubella vaccination at eight months, with days 7-13 relative incidence 1.17 (0.68-2.01). Pooling requires exchangeability across timing, dose history and vaccine strain. These are not precisely the same administration contrast.

Kim et al. (2024) report Korean infant PCV febrile seizures over days 0-7 versus days 15-22, IRR 1.27 (1.10-1.47). PCV13 represents 85.3 percent of the self-controlled analysis; PCV10 and the first three infant doses are also represented. The other three family inputs concern US PCV13 over days 0-1 versus days 14-20. The broader

window/formulation bridge is therefore a sensitivity, not a replacement for the narrower PCV13 estimate of 1.770 (1.116-2.720). The shared US sources are reused across those two families and must not be counted twice.

Each family retains the bias, heterogeneity, sampling-dependence and prior specifications described in S67.5; the four-input PCV family also has four leave-one-out fits. All 37 native families use 48 distinct selected contrasts from the wider published-source census; the chronic-outcome families are described in S69. Repeated family membership does not increase the number of independent studies. The positive PCV signal is retained; neither an autism null-centered estimate nor an inverse estimate on another endpoint establishes universal vaccine safety.

Primary sources: Macartney et al., DOI `10.1016/j.vaccine.2014.10.071`; Xu et al., DOI `10.3390/vaccines9101073`; Kim et al., DOI `10.1093/ije/dyae010`. Source links, risk/control windows, participant identities and all sensitivity fits appear in `literature/all_native_study_dispositions.csv` and `literature/native_family_synthesis.csv`.

S69. Rotavirus and chronic disease associations

S69.1 Sources and comparisons

The chronic-outcome extraction retains 59 contrasts from eight articles and six specified participant-data systems: the U.S. Vaccine Safety Datalink, MarketScan and Optum, UK CPRD Aurum, Finnish national registers and Norwegian national registers. Dose, product, age, alternative adjustment and historical-comparator estimates remain linked to their source populations. They are not 59 independent studies. The source census and extraction locators are in `literature/all_native_study_dispositions.csv`; the pinned transcription rules are in `chronic_study_updates_3402.py`.

Type 1 diabetes contrasts come from Glanz et al. (2020), Burke et al. (2020), Rogers, Basu and Kim (2019), Inns et al. (2021), Vaarala et al. (2017) and Östman et al. (2026b). Coeliac disease uses Inns, Vaarala, Kim et al. (2022) and Östman et al. (2026a). Kim's autoimmune thyroid outcomes are retained separately. The Finnish source is available here only as a primary abstract: its reported relative risks are not silently relabeled as hazard ratios, and full-text dose definitions remain unverified.

The Norway receipt contrasts are observational adjusted hazard ratios, not the interrupted-time-series slope changes described in S49. The coeliac estimate used here is the adjusted Table II HR, 1.02 (0.82-1.28), rather than the unadjusted abstract HR, 0.99 (0.92-1.07). The two Norwegian articles follow the same birth cohort. Rogers and Kim also share an overlapping Optum birth population. Their different outcomes cannot supply independent votes for a universal safety conclusion.

S69.2 Dependence and estimand sensitivity

The analyses use the native-family model and bias assumptions in S67.5. Three full-series HR families pair Norway with one U.S. system at a time. Three any/full-series bridge families additionally admit the UK any-recorded-dose contrast. Three further sensitivities add the Finnish reported relative risk. A tenth type 1 diabetes family admits all three U.S. systems but varies assumed same-country sampling correlation between 0, 0.5 and 0.9. VSD, MarketScan and Optum are not known to be participant-disjoint. Correlation sensitivity is a working dependence model, not an estimate of actual overlap; the one-U.S-system families avoid relying on their disjointness. Two coeliac families respectively retain HRs and add the Finnish RR bridge. All twelve family definitions overlap, so their posteriors are not pooled again.

Table S69-A. Selected chronic-outcome reference summaries.

Outcome and comparison	Selected inputs	Mean ratio	95% credible interval
Type 1 diabetes, full-series HR, Norway and VSD	2	1.093	0.652-1.805
Type 1 diabetes, full-series HR, Norway and MarketScan	2	1.107	0.699-1.738
Type 1 diabetes, full-series HR, Norway and Optum	2	0.874	0.534-1.533
Type 1 diabetes, all-U.S. any/full bridge	5	0.940	0.688-1.301
Coeliac disease, HR family	3	1.032	0.750-1.417
Coeliac disease, Finnish RR/HR bridge	4	1.000	0.751-1.326

The all-U.S. type 1 diabetes summary is 0.942 (0.676-1.330) at sampling correlation 0.5 and 0.941 (0.666-1.352) at 0.9. A larger shared-bias allowance gives 0.944 (0.606-1.478). Omitting the Optum contrast moves the mean to 1.030 (0.744-1.428); omitting other inputs is also reported. These results neither establish prevention nor exclude moderate harm under all plausible confounding models. Follow-up ages, vaccine formulation, receipt definitions and outcome ascertainment differ. No chronic-outcome summary enters the autism policy hierarchy, and no synthesis across type 1 diabetes, coeliac disease and thyroid disease is attempted.

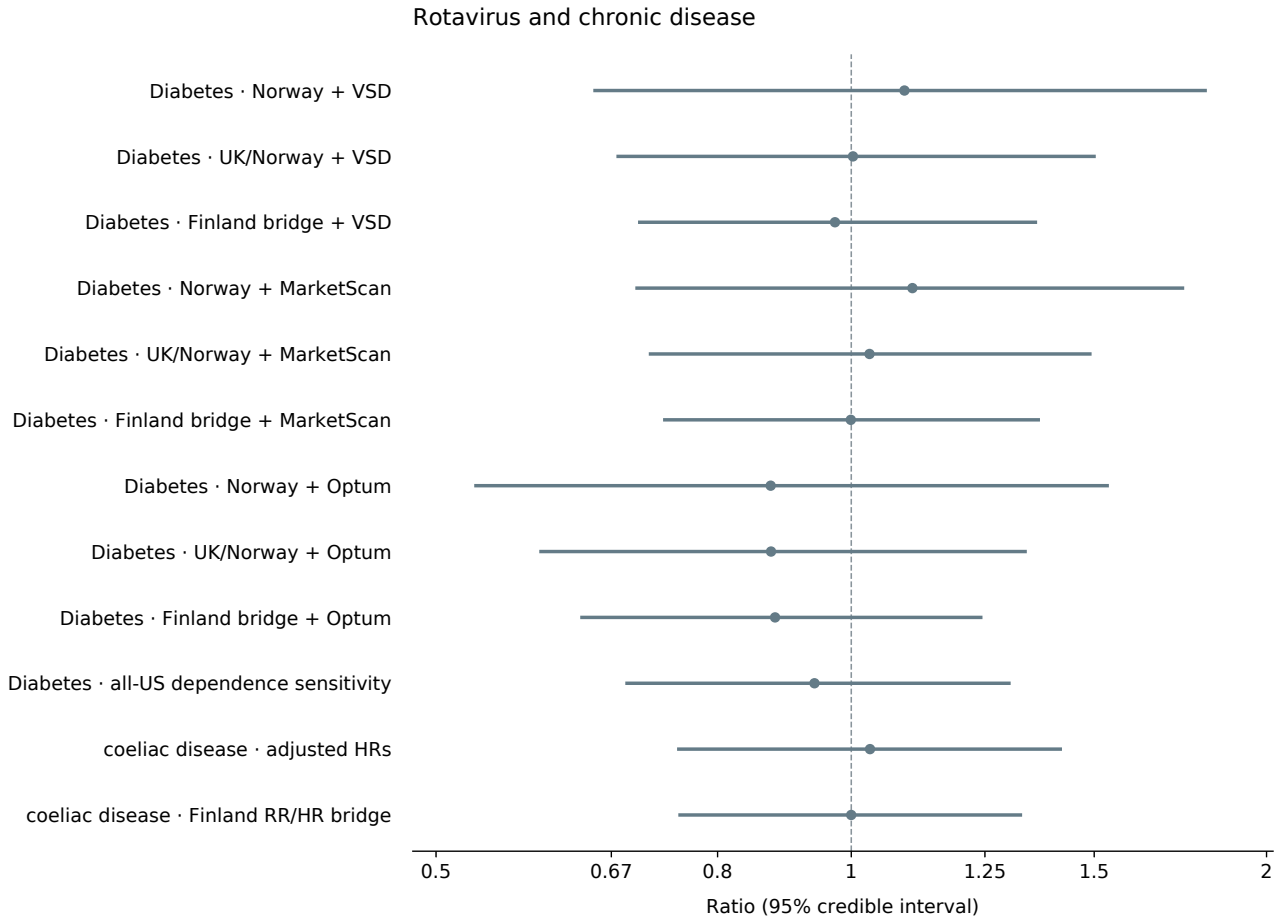


Figure S113. Rotavirus and chronic disease. Points are posterior medians of each family's exponentiated mean log effect; bars are 95 percent credible intervals under the reference bias allowances. Rows are overlapping alternative family definitions, not independent replications. Full-series HR families use Norway and one U.S. system; UK additions bridge any-dose and full-series contrasts; Finland additions also bridge reported RR and HR. The all-U.S. row also includes Norway and the UK, with sampling-correlation sensitivities in S69. Positive and inverse means retain broad uncertainty. These observational summaries do not enter the autism policy hierarchy.

The source archive preserves publisher or repository captures where retrieval succeeded, with URLs, table locators and checksums. The Glanz and Burke publisher tables were checked through the web reader, but direct local retrieval returned access errors; the package does not claim to contain their primary full texts. Primary-source capture and contrast-family tests check the selected values and prevent repeated participant-spine membership within a fit. Reproduction of the numerical model does not validate the observational causal assumptions.

S70. Native units and complete result accounting

The integrated register contains 72,585 records: 60,844 historical records, 4,515 additional published contrast records, 3,158 source-aligned local-model fits, 2,067 explicitly mapped native or diagnostic records and 2,001 derived analysis records, including unavailable model attempts. Records are not independent observations or studies. Every original record is retained with a source-specific disposition; current-value fields distinguish native units, exposure definitions, confidence levels and source values without manufacturing additional study weight. Source-only companions, power scenarios, covariance diagnostics and repeated likelihood calculations have separate tables rather than being counted as independent effect estimates.

The native mapping includes 63 event estimates, 12 Swedish early-rotavirus diagnostics, 26 Scottish school-support estimates, 549 comparative-outcome records, 154 harm-direction display records and 588 regional associations from Ghana, Rwanda, Uganda, Namibia and Ethiopia. Their full source fields and original table hashes accompany each record. Odds ratios per ten percentage points of regional coverage are not individual vaccinated-versus-unvaccinated risk ratios. School-support shares and all-age condition prevalence are not automatically childhood population incidence. Estimates in both directions, unestimated rows and failed diagnostics remain visible.

The held-table audit screened 2,624 effect-like CSV tables. Its residual 143 distinct contents, appearing at 325 paths, have explicit roles and current targets in `accounting/held_source_role_ledger.csv`. The roles distinguish genuine source estimates from copies, older versions, derived posteriors, power calculations, placebo assignments, exposure-only tables and website exports. Numerical differences are linked to primary-source corrections or documented implementation changes. Search-result ranks and data-availability records are not treated as coefficients. Header-based coverage is an accounting check, not a guarantee that every possible dataset or estimator has been exhausted.

The reproduction entry point verifies file integrity, reintegrates supplied likelihood grids and refits all 37 published and seven local reference families from supplied contrast inputs. Source-specific extraction and audit scripts accompany the native mappings and diagnostics. This scoped replay does not reconstruct every historical model from fresh source downloads. Article redistribution rights, access failures, approximate figure recovery and unresolved historical-vintage differences are distinguished from numerical reproducibility.

S71. Calendar alignment and historical cohort diagnostics

S71.1 Local panel timing and inference

Lagged coverage is matched by jurisdiction and exact calendar year: exposure year equals outcome year minus the specified lag. Missing years remain missing. An observation two years earlier cannot stand in for a one-year lag merely because it occupies the previous row. Within the held joined panels, 228 specifications are estimable under this rule; 84 are contemporaneous. Independent fixed-effects projection reproduces the corresponding dummy-variable regressions and cluster-robust CR1 covariance calculations. Coefficient intervals use jurisdiction-cluster degrees of freedom. Coverage is measured per ten percentage points, and outcomes are smoothed disability shares, not individual incident diagnoses.

For example, the Texas HepA lag-one autism-composition ratio is 1.004 (95 percent model interval 0.949-1.061), with 1,763 observations. The lag-two ratio is 0.952 (0.851-1.066). These associations do not identify an individual vaccine effect: kindergarten coverage is not exposure measured for every birth cohort in the multi-age disability denominator. Earlier exposure-only years are not added without a validated geography crosswalk. The downloadable calendar-aligned table reports retained observations, jurisdictions, years, within-unit exposure variation and uncertainty for each specification.

Conditional power calculations cover 304 scenarios: 38 autism specifications, two assumed exposed-cohort shares of the outcome stock and four effect targets. The mixture calculation links an assumed individual effect to a predicted ecological contrast. Its power depends on the exposure first stage, cohort share, outcome ascertainment, model and covariance assumptions. It is not an unconditional individual-risk bound. Fifteen local MMR-HepB interaction coefficients also reproduce under an independent multivariate fixed-effects calculation. Multiplicity adjustment uses the complete stated family, not a selection of small p-values.

S71.2 Montreal schedule transition

Fombonne et al. (2006) report school prevalence across twelve grade cohorts. Birth years are inferred from grade, vaccination coverage comes from a wider region, and removal of thimerosal and addition of a second MMR dose coincide. Grade-specific counts and denominators are not published. Consequently, the individual-data regression cannot be exactly reconstructed from the printed rates and confidence intervals.

The summary-data analysis transforms the reported prevalence and interval endpoints to logits, approximates standard errors from the interval width, and compares weighted and unweighted cohort regressions. Sixty-nine retained result rows cover linear and quadratic trends, segmented trends, schedule-assigned mercury, regional coverage, three candidate boundaries and a six-boundary scan. HC3 and lag-one HAC covariance alternatives use

residual-degree-of-freedom t intervals; only the boundary scan receives its own Holm adjustment. These are sensitivity analyses of one sample, not independent replications or a preregistered model family.

The 1996 linear-trend weighted specification gives a cohort-prevalence odds multiplier of 0.658 (0.403-1.075). It does not establish a protective effect of either schedule change. For the 1997 segmented specification, the two post-boundary observations are exactly fitted and have unit leverage. Six result rows retain their coefficients but have no standard error, interval or p-value: HC3 is undefined, and residual-based HAC cannot recover the innovation variance of those exactly fitted observations. All other nominal intervals remain conditional on the small-sample covariance model.

S71.3 California cohort and calendar series

Dales, Hammer, and Smith (2001) provide fifteen birth-cohort points for autism service-case counts and MMR coverage by 17 and 24 months. Forty-five points are recovered from the original PDF's vector coordinates. The published endpoints are retained separately. Axis-calibrated endpoints differ from printed endpoints by up to six cases and 1.1 coverage percentage points, so the analysis includes both axis calibration and a printed-endpoint anchoring sensitivity. Neither is presented as exact recovery of unpublished observations.

Sixteen descriptive specifications cross the two calibrations, two vaccination-age checkpoints, linear or quadratic cohort trends, and HC3 or lag-one HAC inference. For the printed-endpoint calibration and 24-month coverage, the linear-trend count ratio per ten coverage points is 1.124 (0.836-1.512); the quadratic-trend ratio is 1.032 (0.847-1.258). Annual birth denominators are unavailable, attained age varies by cohort, and vaccination and service records are unlinked. The nominal intervals do not include plotting uncertainty. These estimates therefore remain service-count associations.

Schechter and Grether (2008) report 2004-2007 quarterly prevalence slopes among California DDS clients aged three to five: 0.09 (0.08-0.09) autism cases and 0.04 (0.03-0.05) total developmental-disability cases per 1,000 live births per quarter. The rounded intervals are retained as printed. Total disability includes autism, so it is neither an independent outcome nor automatically a negative control. These slopes do not estimate a withdrawal discontinuity or an individual thimerosal effect. The California analyses share a reporting system and receive no additional independent weight in the main synthesis.

S71.4 Numerical reproducibility and limits

The portable archive provides separate commands for file-integrity verification, saved-likelihood reintegration, all 44 native reference-family fits, and the Montreal and Dales cohort diagnostics. The last command refits supplied rounded or figure-derived observations; it does not reacquire the primary articles. Full-project audit scripts additionally reproduce 58 state specification-curve rows and the state two-way wild-bootstrap result. Numerical agreement validates those implementations and inputs, not their identifying assumptions. Source access failures and unresolved historical-input vintages remain listed in the acquisition and provenance ledgers; unavailable information is never coded as a null effect.

S71.5 Archived school panels and short service-area comparisons

Table S71-A. Source-aligned local and district sensitivity families. Counts describe overlapping attempts or fits, not independent causal estimates.

Family	Attempted specifications	Estimated	Unavailable	Native target
Oregon archived school coverage	340	242	98	District autism identification rate and other reported disability outcomes
Texas county coverage, available-year rule	32	32	0	Disability composition among paired reporting LEAs
Texas complete three-year exposure window	28	28	0	Fourteen native fits and fourteen matched-support fits on the same composition outcomes

Family	Attempted specifications	Estimated	Unavailable	Native target
Oregon preschool service areas	18	16	2	Autism share of reported EI/ECSE special-education counts
South African district health services	226	214	12	Mental-health contacts and native PHC treatment measures

Oregon's archived school comparison uses five school-frame and vintage scenarios. Internally labeled 2022-23, 2023-24 and 2024-25 waves form the primary dated sample; capture-inferred 2017-18, 2019-20 and 2021-22 waves are sensitivities. Balanced-frame alternatives distinguish changes in coverage from changes in which schools report. Suppressed counts retain lower, midpoint and upper alternatives. The 340 attempts comprise 260 lag specifications, 40 other-outcome fits and 40 coverage-pair specifications; respectively 182, 40 and 20 are estimable. Future-wave pair diagnostics are labeled by their actual calendar gap, not presumed to be one-year leads. The 136 available power scenarios use normalized conditional exposure variation and the native ecological coefficient. Whole-endpoint alternatives are measurement sensitivities, not sharp causal bounds; descriptive rate and odds transformations are not individual vaccination effects.

The Texas available-year rule averages contextual kindergarten coverage over outcome year y through y+2 when at least one feed year is observed. The complete-window sensitivity instead requires all three exact years, without substituting zeros for missing years. Recovering the 2021 kindergarten county source expands coverage support; it does not add an independent outcome population. For early-childhood autism's share of all early-childhood disability, the complete-window baseline ratio is 0.752 (0.330-1.712; 359 county-years, 205 counties), compared with 0.838 (0.606-1.160; 703 county-years, 226 counties) using the recovered source. Both are per ten percentage points of contextual coverage. The common-support exercise holds 499 county-year exposure keys fixed; coefficient differences between the paired fits have no separately estimated covariance or difference standard error. Neither sample expansion nor a smaller interval establishes causal identification.

Oregon preschool exposures measure coverage among listed sites, not all resident children or the EI/ECSE recipients whose disability composition is observed. Nine service areas and two or three waves permit only short ecological comparisons. Eighteen contemporaneous attempts retain sixteen coefficients and two unavailable fits; 54 exact-calendar lag-feasibility records and 72 leave-one-area diagnostics remain separate. Holm adjustment includes all nine antigen attempts in each vintage family, assigning unavailable attempts p=1 solely for family accounting. MMR with an unspecified dose is not relabeled MMR1. The ten-point coefficient is a rescaled slope, often many times larger than observed conditional exposure variation, and cannot support extrapolation to a ten-point vaccination intervention. CR1 inference with t(8) is a small-cluster convention, not verified coverage.

South Africa's district series distinguish under-18 mental-health visit counts from treatment percentages: contacts are not unique diagnosed children, and no population denominator is inferred. Coverage above 100% remains flagged administrative coverage, not a literal exposure probability. The eighteen primary comparisons use district and financial-year two-way clustered t inference with only eight to ten financial-year clusters. One primary coefficient is nominally significant, but none survives BH or Holm correction in either the eighteen-primary or 214-finite-fit family. Twelve two-year specifications remain unavailable. The temporal and scale alternatives are dependent service-use diagnostics, not autism estimates or evidence of causal protection.

Sources: `qa/data_handling_20261002/oregon_archived_refits_3557/archive_run_005/receipt.json` and its estimate/power tables; `texas_county_refit_3567/run_20261005T061245615456Z/paired_effect_estimates.csv`; `texas_complete_window_3593/run_20261005T064827694831Z/complete_window_native_estimates.csv` and common-support tables; `oregon_preschool_screens_3600/run_002/receipt.json`; and `south_africa_dhb_3582/finalization_summary.json` and `finalized_screens_multiplicity.csv`, all latter paths under `qa/data_handling_20261002/`. These families reuse existing populations and add no independent weight to the 17-context autism hierarchy.

S71.6 Early-childhood composition across six states

An explicitly descriptive synthesis uses one contextual kindergarten-MMR coefficient from Connecticut, Kansas, Kentucky, Ohio, Pennsylvania and Texas. The outcome is autism identifications divided by all early-childhood disability identifications among paired reporting LEAs. Connecticut's exposure is LEA-level; the other exposures are county contextual proxies. The retained y-through-y+2 rule requires at least one reported feed year, so this is not a uniformly complete-window exposure analysis across states.

The six-state mean ratio is 1.002 per ten coverage percentage points. Its modified Hartung-Knapp interval is 0.761-1.321, using five degrees of freedom and a variance multiplier no smaller than one. The conventional DerSimonian-Laird normal interval is 0.812-1.237; the unmodified Hartung-Knapp interval is 0.774-1.298. Forty-five interval and leave-one-state summaries and eighteen fixed-weight covariance sensitivities remain visible. Assumed sampling correlations of 0, 0.25 or 0.5 and shared-bias SDs of 0, 0.10 or 0.20 on the log scale are sensitivity assumptions, not measured dependence. Disjoint states may still share reporting or diagnostic shocks. The two Texas exposure vintages enter alternate pools, never as two independent states. This family is distinct from the twelve-state school-age MMR composition family in S67.3. Neither composition outcome supplies population autism incidence, an individual vaccine-risk ratio or additional autism-hierarchy likelihood weight. Source: `qa/data_handling_20261002/composition_six_state_3592/run_001/pool_results.csv`, `members.csv` and `covariance_sensitivity.csv`.

S71.7 Dependent age and cohort extensions

South Carolina age/cohort models and federal IDEA ages-18-21 models supply 32 dependent specification attempts: 30 coefficients and two unavailable fits. The 1,120 exposure-only refits are sensitivity diagnostics, not additional effect estimates. South Carolina's resident-rate and IEP-composition models share autism numerators; the older-band models share observed band totals. They also overlap existing cohorts, NIS household draws and Census inputs. Exponentiated ecological slopes per ten percentage points are not individual vaccine risk ratios.

All 30 estimated specifications have failed or non-positive-semidefinite primary two-way covariance diagnostics, so their primary confidence intervals remain unavailable. The two rank-deficient older-band specifications have unavailable coefficients, not zero effects. State-only intervals and raw target-variance diagnostics remain explicitly qualified and do not replace valid joint inference. A separate calculation reproduces South Carolina's 16 point coefficients, but shares its source/design loaders and does not validate causal identification or uncertainty coverage. Sources: `qa/data_handling_20261002/dependent_age_extensions_3679/REPORT.md` and `compact_run_002/registered_native_records.csv`.

S71.8 New York age-band counts and reporting regimes

The historical district age-band sensitivity analysis evaluates 144 model attempts across six zero-handling policies, four source regimes, three outcome series and two calendar knots. There are 119 fits, 17 rank failures and eight insufficient-sample attempts. Each attempt supplies a level-change and a slope-change record: 288 parameter records, not 288 models or independent events. Literal source zeros may represent suppressed counts. Generic allocations of zero, two or four are disclosure stress scenarios, not observed values; conditional marginal allocations respect the displayed totals where feasible. Dropping ambiguous zeros changes sample selection. Negative source residuals and suppression exceptions remain flagged, not clipped into feasible values.

Table S71-B. Autism counts at ages 4-5 under the literal-zero policy. Entries are changes in log conditional reported count per calendar year, not vaccine-risk ratios. Calendar terms include the 2007 file-format boundary, the October-2008 count-date boundary and the 2012 reporting boundary.

Source regime	2006 knot	2007 knot
Continuous 2003-2012	0.130062	0.133871
Through 2011	-0.021737	-0.067274
All years, separate 2012 level	-0.021651	-0.067272
All years, source/calendar level terms	-0.055634	Rank failure

The sign reversal shows that the continuous-series slope change is not robust to the reporting boundary; it does not prove that the nuisance terms explain or repair the unknown change in reporting universe. The literal-zero continuous

series has 2,073 informative observations from 208 districts at ages 4-5 and 4,469 from 448 districts at ages 6-11. The historical four-category comparator has only 150 observations from 15 districts and is not assumed vaccine-unaffected. County-district labels are not verified longitudinal identifiers. The models have no population offset, measured vaccine first stage or exact exposed-cohort assignment: growth in reported counts is not incidence, prevalence or an individual RR.

Full two-way covariance matrices are non-positive-semidefinite for 111 of 119 fits. The other eight have only nine or ten calendar-year clusters and supply nominal uncertainty, not validated causal intervals. Independent calculation reproduces two selected 2,073-row fits and checks the reported interval arithmetic across all 119 fits; it does not independently refit every covariance matrix or establish statistical coverage. These dependent diagnostics receive no independent synthesis weight. Source-cell and model receipts: `qa/data_handling_20261002/ny_source_cell_reconciliation_3681/receipt_final_3681.json`, `ny_band_tracking_sensitivity_3687/run_002/REVIEW.md` and `nominal_estimates.csv`; independent review: `ny_sensitivity_review_3688/RUN_002_ADDENDUM.md`, all paths under `qa/data_handling_20261002/`.

S71.9 Historical ages 18-21 classification series

The source-aligned state/DC bridge contains 1,102 numeric autism-classification counts out of 1,122 possible observations in 2000-2021. The historical IDEA Part B table explicitly concerns students receiving services, not clinical adult prevalence. Direct ages-18-21 totals are retained when their single-age components are incomplete; no single-age autism counts are fabricated. The series crosses reporting-source boundaries in 2005 and 2012 and resident-population vintage boundaries in 2010 and 2020. School retention, graduation, eligibility and classification can change the numerator; population weights cannot recover the cases' actual age or vaccine-exposure composition.

The available-state classification/resident ratio rises from 2.80 per 10,000 in 2000 to 22.52 in 2015 and 35.54 in 2021. Both 2000 and 2015 include all 51 jurisdictions; 2021 includes 50 and is not a complete national rate. July-1 resident denominators are not literal IDEA count-date populations. October-1 and December-1 reference dates are sensitivity assumptions, not verified historical dates.

Table S71-C. Eight dependent count-model sensitivities, each using 1,102 observations, 51 jurisdictions and 22 years. Ratios are exponentiated ecological coefficients per ten percentage points in the named construct. Primary joint confidence intervals are unavailable for all eight models.

Exposure construct	Assumed date	State/year effects	Plus state trends
Infant structural-policy scenario	December 1	1.024349	0.993914
Infant structural-policy scenario	October 1	1.027124	0.994219
Post-2000 eligibility share	December 1	1.191646	1.053150
Post-2000 eligibility share	October 1	1.245113	1.038030

The infant construct uses first pneumococcal dose before 213 days, not verified pure PCV7 receipt. Birth windows through 1998 receive an assumed structural policy zero; this is not observed individual non-exposure, and trials, catch-up and other pneumococcal use qualify its interpretation. Eligibility share is not observed uptake or an instrument. The eight fits condition on fixed held exposure points without propagating first-stage sampling uncertainty. All full state/year covariance matrices are genuinely non-positive-semidefinite; a positive focal variance does not repair the joint matrix. State-only and raw focal intervals remain diagnostics, not valid replacements. An independent implementation reproduces all eight coefficients and covariance calculations conditional on these inputs and assumptions. These results neither add independent rollout evidence nor supply individual-risk or equivalence bounds. Sources: `qa/data_handling_20261002/older_age_historical_feasibility_3689/REPORT.md`, `older_age_long_panel_3694/REPORT.md` and its model/descriptive tables, and `older_historical_numeric_review_3697/REPORT.md`.

S72. Aluminum contrasts and partial-source evidence

S72.1 One cohort, several estimands

The corrected supplementary material for Andersson et al. (2025) supplies 547 contrast entries in Figures 3-12: 517 numerical estimates and 30 explicitly not-estimable entries. The source file is the corrected 21-page supplement archived on 24 July 2025. Every numerical triple and unavailable cell has been compared with the rendered source tables. The register retains author-reported rounding, confidence intervals and comparison direction. These entries reuse the cohort of 1,224,176 Danish children, including overlapping endpoints and nested subgroups; they are not 547 independent studies or additional policy-event likelihoods.

Figures 3-8 and 12 report Cox hazard ratios per milligram of aluminum, varying sex, birth period, inclusion of zero-exposure children, exposure and follow-up starting age, follow-up ending age, and complete-case selection. Figure 9 compares exposure above and below the median. Figure 10 compares low (>0-1.5 mg) and middle (>1.5-3 mg) exposure with high (>3-4.5 mg) exposure. Figure 11 reports age-five risk differences for those two comparisons, estimated separately using stabilized inverse-propensity-weighted Kaplan-Meier curves. Covariates include calendar year, season, sex, birth characteristics, early healthcare contacts, maternal characteristics and household income. The source says covariate balance was assessed but does not display it. Overlap, weighting diagnostics and the joint sampling covariance cannot be independently recovered from these aggregate tables.

Table S72-A. Selected autism contrasts illustrating the difference between measures. Each interval is the source's nominal 95 percent interval; the high-minus-middle row is an exact sign reversal, not a separate estimate.

Contrast	Measure	Estimate	95% interval
Per mg, follow-up ages 2-5	HR	0.93	0.89-0.97
Per mg, follow-up ages 2-8	HR	0.95	0.92-0.97
Low versus high dose	HR	1.29	1.11-1.50
Middle versus high dose	HR	0.97	0.90-1.05
Low minus high dose	RD per 10,000	13.09	-2.20 to 28.39
Middle minus high dose	RD per 10,000	-8.68	-12.38 to -4.98
High minus middle, re-expression	RD per 10,000	8.68	4.98-12.38

The positive high-minus-middle risk difference must not be described as a null finding. Conversely, the positive low-versus-high hazard ratio must not be presented as evidence that more aluminum increases risk. The hazard and risk-difference models differ in scale, adjustment method and target; one cannot select the preferred answer or algebraically translate one into the other. Neither pattern establishes a causal dose response under the available evidence. Source-specific confounding, calendar/formulation variation, weighting assumptions, rare events and correlated testing remain relevant. An inverse hazard ratio is not proof of protection.

The birth-period analysis includes an Asperger-syndrome HR of 1.67 (1.01-2.77) in children born 2007-2018, based on 51 events, and an idiopathic thrombocytopenic purpura HR of 1.37 (1.07-1.76) in 1997-2006 births. All positive as well as inverse results remain in the downloadable census. Across the 517 numerical entries, 20 nominal intervals lie wholly above their native null and 129 wholly below; the null is one for HR and zero for RD. These counts mix comparison directions and strongly dependent specifications. They are not counts of established harms, protective effects, independent replications or false positives, and no multiplicity-adjusted inference is claimed from the rounded intervals.

S72.2 Source consistency and reproducibility

Figure 6's caption and methods define exposure and follow-up from 14 months although its column header repeats age two; the caption governs the recorded age. Figure 11's caption says risk differences above 1.00, but the RD null is zero. Figure 7 presents ages 2-5, 2-4 and 2-3 in that order; the main-model column repeats existing evidence. Thirty unavailable entries remain missing, never zero. The structured extraction records the original row text, exact page, column, source hash, exposure age and follow-up age. Label checks flag effect rows with unrecognized disease names, preventing silent OCR omissions. Native RD entries cannot enter a log-ratio synthesis.

The preprint by Jablonowski and Hooker (2025; DOI 10.20944/preprints202508.0280.v1) reuses this Danish cohort rather than adding a sample. Its reversed risk differences and positive Asperger subgroup are represented here from the corrected primary source. Its crude cohort-denominator calculations do not reconstruct the adjusted time-to-event analyses; inverse hazard-ratio percentages are not ordinary risk reductions. Source disagreements do not, by themselves, establish the preprint's allegations of fraud. The article's conclusions and critical commentary are assessed against the same retained primary numbers.

S72.3 Partial abstracts and non-vaccine contrasts

Seven other recoverable primary-abstract intervals remain explicitly partial extractions. Chen et al. (2004; PMID 15259839) report a 21 percent reduction (6.9-33.3 percent) in sampled autistic-disorder births after 1987, equivalent to a cohort-period ratio 0.790 (0.667-0.931), not an individual MMR effect. Geier et al. (2007; PMID 17674242) report maternal Rh negativity OR 2.35 (1.17-4.52) against nonconcurrent prenatal-clinic controls; the blood-group proxy does not measure childhood vaccine receipt. Schultz et al. (2008; PMID 18445737) report three positive, imprecise acetaminophen associations after MMR, ORs 6.11, 3.97 and 8.23, in overlapping subgroups of a self-selected parent survey. The exposure is a drug and the groups are not independent studies.

A Danish HPV referral study (PMID 30803841) reports prevalence-proportion ratios 1.24 (1.21-1.27) for any hospital contact and 1.97 (1.76-2.19) for at least six contacts during the five years before vaccination. These pre-exposure contacts cannot be effects of the subsequent vaccination; neither do they establish that later symptoms are psychogenic. The 1,496 referred cases and 7,480 controls overlap the earlier referral study (PMID 28979164). Full-text endpoint and subgroup inventories for the abstract-only sources remain acquisition tasks. No missing estimate is entered as a null or as an independent vaccine-risk likelihood.

S72.4 Finnish reported-event follow-up

The original Peltola et al. (1998) letter, reproduced on archive PDF pages 285-286 of the ERIC hearing collection, follows 31 reported gastrointestinal cases after MMR vaccination. None developed an autistic-spectrum disorder during reported follow-up averaging nine years and three months. The source reports approximately three million vaccine doses delivered during the program; that is not a count of people actively followed and found autism-free. Twenty of the 31 reported cases were hospitalized and five experienced febrile seizures. Reporting was incomplete by the authors' acknowledgment. The case series supplies descriptive symptom follow-up, not a comparative vaccine-autism risk estimate, and no binomial population-safety bound is manufactured from the dose total.

S72.5 Danish thimerosal withdrawal and registration changes

Madsen et al. (2003) report 956 first autism registrations among children aged two to under ten during 1971-2000. Their three attained-age curves show calendar incidence continuing to rise after the 1992 thimerosal phase-out. The source describes a historical pertussis schedule at five weeks, nine weeks and ten months, totaling 125 micrograms of ethylmercury. March 1992 is the last batch-release date reported in that article, not an established last administration date; the phase-out was gradual.

The outcome system also changed: ICD-10 replaced ICD-8 in 1994, and outpatient contacts were added in 1995. The authors report a similar upward inpatient-only trend but do not provide that series. These changes are part of the explanation of recorded incidence, not quantities that can be attributed to vaccination. Differencing the three age curves would require assumptions about age-specific ascertainment and cohort trends; the older children are not automatically an unaffected control group. The article supplies neither an adjusted vaccine-risk coefficient nor exact annual counts and denominators. We retain its printed descriptive facts and the primary figure without constructing a precise withdrawal RR from the raster plot. It contributes no additional independent Danish likelihood.

S72.6 Neuropsychological sensitivity and effect-modification models

The official CDC technical report accompanying Thompson et al. (2007) provides 878 additional coefficients in Exhibits 9.2.2.4, 9.2.5.2 and 9.2.6.1-2. Eighty examine ADHD-rating outcomes after excluding medicated children or treating medication use or a high symptom score as a binary outcome. The prose says 38 medicated children whereas the table says 39 excluded; model-specific sample sizes are therefore left unspecified. Forty repeated main-model coefficients are cross-checked but not added again.

Another 210 coefficients concern a composite prenatal exposure-source score and its interaction with infant ethylmercury exposure. The score combines six vaccine, immune-globulin and environmental-source indicators; it is neither a mercury dose nor a vaccine count. The remaining 588 coefficients model exposure with and without concurrent antibiotics, antibiotic-treatment days, effect modification and reported sums of exposure coefficients. Concurrent exposure includes antibiotic courses and the 14 days on either side. Reported standard errors for coefficient sums are used directly, not reconstructed by assuming zero covariance.

There are 33 printed nominal p-values below 0.05 among these 878 overlapping coefficients. Approximate normal-Wald intervals reconstructed from rounded coefficients and standard errors exclude their native null for 32; the difference reflects rounding and source test conventions. These are not multiplicity-adjusted discoveries. For example, the assessor-rated motor-tic coefficient for concurrent exposure during days 1-214 is 0.084 (SE 0.040), corresponding to an OR of 1.088 (calculated interval 1.006-1.176) per microgram/kg. This association is retained alongside estimates of the other model terms; it is not a causal vaccine-injury estimate. Illness-related antibiotic use, selective enrollment, model selection and repeated testing complicate interpretation.

These models reuse the 1,047-child cohort and did not assess autism; they add no participant weight to the autism synthesis. The census preserves each coefficient, unit, page and model term. Appendix C's influence analyses and Appendix E's power calculations are described below. Appendix D's coefficient paths follow a hybrid covariate-selection procedure and require separate treatment.

Table S72-B. Sex-subset models without a prenatal-by-infant exposure interaction in Volume I of the CDC technical report, Exhibits 9.2.4.9b and 9.2.4.10b, printed pp.150-151/PDF pp.158-159. These are continuous-outcome coefficients, not ORs or RRs. Prenatal exposure is per microgram of ethylmercury; infant exposure is per microgram/kg accumulated during days 1-214.

Outcome and subgroup	Exposure	Coefficient	SE	Approximate 95% interval
Long-delay verbal recall, female	Prenatal	-0.010	0.013	-0.03548 to 0.01548
Long-delay verbal recall, female	Infant	-0.022	0.017	-0.05532 to 0.01132
Teacher-rated hyperactivity/impulsivity, male	Prenatal	0.036	0.034	-0.03064 to 0.10264
Teacher-rated hyperactivity/impulsivity, male	Infant	-0.058	0.056	-0.16776 to 0.05176

Intervals are normal-Wald calculations from rounded source coefficients and standard errors, not printed source intervals or multiplicity-adjusted inference. Model-specific sample sizes and degrees of freedom are unavailable. The 538 female and 509 male participants are subsets of the same 1,047-person cohort, not new samples. Ninety exposure metadata cells represent 45 repeated sex/exposure slots; 936 participants have zero recorded prenatal exposure and only 13 exceed 25 micrograms, limiting interpretation of sparse joint-exposure patterns. No autism endpoint or independent synthesis weight is added. Source tables and execution pins: `qa/data_handling_20261002/thompson_appendix_gaps_3692/v2/run_002/summary.json` and `subset_no_interaction_coefficients.csv`.

S72.7 Influential observations, exposure coding and reported power

Volume II of the CDC technical report, Appendix C, examines whether extreme exposures or outcomes create or obscure associations. Exhibits B.43-B.44 provide 252 sex-specific coefficients for 42 outcomes. Thirteen prenatal exposures above 25 micrograms are capped at 25, neonatal exposure becomes an any-versus-none indicator, and the days 29-214 exposure term retains its dose scale. The neonatal indicator includes vaccine and immune-globulin sources. These are changed exposure contrasts, not simple unit conversions of the continuous-dose model. Capping and dichotomization can also attenuate a genuine dose response.

Two further coefficients exclude one boy whose extreme finger-tapping performance was verified in the records; his exposure was not unusually high. The neonatal coefficients change from 0.41 to 0.37 for dominant-hand tapping and from 0.27 to 0.25 for non-dominant-hand tapping. The report gives standard errors of 0.13 and 0.12, and p-values of 0.006 and 0.046, for the leave-one-out results. The reported p-values are retained rather than overwritten by calculations from rounded coefficients.

The 254 sensitivity coefficients include 19 nominal p-values below 0.05, with both favorable and adverse directions. A source inconsistency matters: the narrative calls the adverse female language result speeded naming, but Exhibit B.43

reports it for comprehension of instructions (-0.575, SE 0.248, p=0.021). Speeded naming is -0.130 (SE 0.692, p=0.851). Female verbal IQ is -3.496 (SE 1.202, p=0.004) for the any-versus-none neonatal contrast. These observational, overlapping comparisons are retained without treating nominal significance as causal evidence or independent replication.

Appendix E, Exhibit E.2, gives 105 power calculations for 35 continuous outcomes and three exposure terms. For a stipulated standardized regression coefficient of 0.10, reported power ranges from 67% to a rounded 100%; 99 of 105 entries report at least 80% and 66 report at least 90%. These are source calculations using the fitted design and its standard errors, not power against an individual autism risk ratio. The reported 80%-power minimum detectable standardized effects range from 0.054 to 0.117; the 90%-power range is 0.063 to 0.135. The prose's lower endpoints (0.056 and 0.65) conflict with the table. All 105 table entries satisfy rounding-aware checks against the report's stated formulas. No power record is counted as an effect estimate, and none supplies additional participants or an autism likelihood.

S72.8 Covariate-selection paths and source discrepancies

Appendix D contains 126 coefficient paths, comprising 4,587 plotted estimates and nominal confidence intervals for the same cohort. The extraction uses the PDF's vector marks, numerical axis labels and covariate labels. Coordinates, source pages and digitization tolerances accompany every point. All intervals are centered on their plotted coefficients within coordinate tolerance. These are recovered source diagnostics, not new fitted models or independent studies.

The source describes hybrid selection: covariates can be retained a priori, by a p-value threshold of 0.20, or when removing them changes an exposure coefficient by more than ten percent. Retained variables remain in subsequent models. A dotted divider separates omitted variables from those retained in the final model; points to its right probe retained covariates. The report does not supply each intermediate design matrix. Neither uniform cumulative deletion nor uniform one-at-a-time omission from the final model is assumed. Source intervals do not account for uncertainty from selecting the model. Tics and stuttering also required forward-selection steps after quasi-separation prevented full-model fitting.

Ninety-four paths have compatible final coefficients and interval widths when allowing vector-digitization error and the main summary table's rounding. Twenty-six further paths have unresolved numerical discrepancies despite agreeing outcome and exposure labels; three have no displayed final-model point, and three have conflicting labels. These categories are mutually exclusive. Five of the 123 displayed final intervals have widths incompatible with the main-table coefficient standard errors under the recorded comparison; all five also have coefficient discrepancies. Missing final points are not null estimates.

The conflicts are visible in the source. The speeded-naming plot captioned Exp07mos repeats the prenatal curve and labels its axis PreNatThimer. The non-dominant pegboard plot in that section has an internal dominant-hand title. The HepB parent-metacognition plot has an internal behavior-regulation title. No label is silently repaired. For the six verbal, performance and full-scale IQ Model 1 paths, the retained-covariate names match the detailed main regression tables, so an obviously different named covariate set does not explain the coefficient discrepancies. The source specifies single imputation for both these diagnostics and the main models; multiple imputation is a separate sensitivity table. Unreported transformations or model versions cannot be reconstructed from the report.

All paths remain available in the diagnostic companion tables, including discrepant and positive estimates. They do not replace the main coefficients, enter an autism likelihood, or constitute a count of successful robustness tests. Nine selected source pages were visually checked; vector and numerical checks cover the full extracted collection, without a claim of complete visual certification.

S72.9 Quantitative extraction gaps and narrative results

Several screened sources lack complete quantitative extraction. Their indexed primary abstracts and available partial captures are retained with separate acquisition, extraction and synthesis statuses. Missing numerical results are not assigned a null or an interval. A valid PDF file or a matched citation does not establish that the full article or its results tables were acquired. The quantitative companion for MMRV seizures below is distinguished from its still-incomplete journal-article extraction.

Young et al. (2008) and Geier et al. (2017) report positive associations. The held PDFs contain one page each, not the complete nine- and seven-page published articles. Young's exposure and outcome prevalence are aggregated by birth

cohort despite originating from 278,624 individual records. Geier's abstract describes 164 atypical-autism cases and 15,216 controls with overlapping exposure windows. Their positive findings remain explicit; neither their absence from a quantitative pool nor incomplete access is evidence of a null effect.

The two VSD studies also share overlapping birth years; participant overlap is plausible but not verified. Uchiyama et al. (2007) examines regression among 904 clinic patients already diagnosed with ASD, not population autism incidence. Its abstract reports no association but gives no effect or interval. Stehr-Green et al. (2003) compares ecological exposure and ascertainment trends across California, Sweden and Denmark; its abstract supplies no quantitative coefficient or interval. Their full tables remain unavailable in the held captures. These gaps cannot establish numerical nulls, individual risks or independent samples. Source dispositions and acquisition pins are supplied in `qa/data_handling_20261002/literature_autism_gaps_3690/REPORT.md` and `source_pins.json`.

Croen et al. (2008) compares 400 autism cases and 410 controls. Its rounded prenatal anti-D exposure percentages (10.0% and 9.3%) do not supply the adjusted coefficient or its interval. Anti-D immune globulin is not a vaccine. No adjusted autism effect is reconstructed from these margins or from assumed complete-case denominators.

Kawai et al. (2014) reports no elevated risks after inactivated influenza vaccination but omits the numerical estimates from its abstract. For live attenuated influenza vaccine, fewer than three events occurred for each monitored outcome and the authors explicitly could not estimate relative risks. Sparse events are not zero events or RR=1. The approximately 3.6 million inactivated and 250,000 live attenuated first doses are not an independent-participant total across VSD publications.

Tartof et al. (2014) studies 3,348 children who already had a febrile seizure: 383 vaccine-associated and 2,965 non-vaccine-associated index events. Subsequent outcomes concern this selected population, not epilepsy incidence among vaccinated versus unvaccinated children. Its combined counts of admissions, recurrences and epilepsy cannot recover the missing exposure-specific adjusted comparisons.

Cook et al. (2019) reports a positive MMRV-seizure association without numerical abstract estimates. The 2013 FDA/Mini-Sentinel companion quantifies seizures during days 7-10 after MMRV versus separate MMR+V at ages 12-23 months.

Table S72-C. Overlapping safety analyses with method-specific samples; source confidence intervals are unavailable.

Method / stopping look	MMRV events / doses	MMR+V events / doses	Adjusted comparison
IPTW, look 4 / day 637	18 / 17,321	7 / 17,502	RD 5.2 per 10,000; RR 2.86
GEE, look 8 / day 1,001	45 / 48,233	13 / 35,137	RD 5.3 per 10,000; OR 2.37

Adjusted rates are 0.080% versus 0.028% (IPTW) and 0.091% versus 0.038% (GEE). Locators are Table 9, printed p.14/PDF p.17; Table 20, printed p.31/PDF p.34; and comparison, printed pp.35-36/PDF pp.37-38. Sequential statistics cannot recover missing confidence intervals. These are recognized seizure harms, not autism effects. Identity with the journal cohort remains unconfirmed. Held verification is a text capture, not a locally hashed or visually reviewed PDF: `qa/data_handling_20261002/literature_safety_gaps_3691/companion_source_capture.json` and `parent_text_verification_receipt.json`.

S72.10 Parameter boundaries in aggregate-to-individual conversions

An algebraic conversion must respect the parameter space of its exposure model. For ordered exposure probabilities, a positive affected baseline-risk share and a nonnegative individual risk ratio, the homogeneous mixture gives

$$Q(R) = \frac{1 + wp_1(R-1)}{1 + wp_0(R-1)}, \qquad 0 \le p_0 < p_1 \le 1, \quad 0 < w \le 1, \quad R \ge 0.$$

The smallest compatible aggregate ratio and the upper limit for positive baseline exposure are

$$Q_{\min} = \frac{1 - wp_1}{1 - wp_0}, \qquad \lim_{R \to \infty} Q(R) = \frac{p_1}{p_0} \quad (p_0 > 0).$$

An aggregate point estimate outside this range has no admissible finite inverse. Clipping its algebraic inverse to zero does not establish complete protection. Confidence sets are instead intersected with the model's feasible aggregate range before inversion. A valid interval can include R = 0 even when the unconstrained point estimate has no admissible inverse; an empty intersection is neither a null finding nor a zero-width interval.

The historical-workbook companion audits 60 pooled-gradient scenarios and 20 design-specific conversions against 540 source values and 340 cached formulas. Fifteen historical pooled point values were clipped to zero despite lying outside the mixture model; the companion records these points as unavailable. Thirty-nine lower confidence boundaries legitimately reach zero under the stated algebra. Two spreadsheet check formulas subtract a blank point cell as though it were zero; neither check is treated as evidence of a valid point estimate. The workbook and historical source versions are preserved, with current values and mapping restrictions recorded separately. None of these overlapping scenarios adds independent likelihood weight or reinstates an invalid exposure-to-risk interpretation.

Odds-ratio conversion likewise distinguishes missing values from unbounded endpoints. A missing RR has no OR. Eighteen point-OR fields across six current scenarios are unavailable for this reason; the RR values, finite OR values and interval endpoints are unchanged. Rebuilding the conversion table from its design inputs and testing positive, negative, null and boundary cases verifies these numerical distinctions. The combined register links its 20 historical conversion scenarios to five parent estimates, preserves the archived aggregate values and supplies current conversion values and restrictions separately. None is an additional independent observation. These checks do not establish the causal assumptions of the mixture model.

S72.11 Suppression flags and derived denominators

Serialized Boolean flags are decoded explicitly as true/false or 1/0. Missing or unrecognized tokens cause an error rather than determining inclusion by ordinary string truth testing. Failure-injection tests verify that equivalent text and Boolean flags preserve the Canadian analysis samples and UK denominator-provenance labels. Tests also reject changed rates, denominators, case counts, suppression flags and duplicate source keys.

The held Canadian CCDSS exports match all 2,016 province/sex/year/measure records in the processed table, including 6,048 rate, population and rounded-case comparisons with missingness preserved. All 608 rows marked suppressed in the processed table correspond to blank source rate fields; this missing-value flag does not distinguish statistical suppression from other source unavailability. No unexplained nonnumeric rate token is coerced to zero. The staggered-program panel reproduces its 382 rows. Synthetic-control input matrices retain 190 observed outcome cells for both sexes, 190 for males and 175 for females across eight provinces and 24 fiscal years. These are overlapping surveillance cells, not independent samples. The current flag checks do not change their values or inclusion.

The UK population cache contains 790 reported cells and no derived-population rows. The proposed identity UK = England and Wales + Scotland + Northern Ireland fails across the 140 comparable 2001–2020 age/year cells, with a maximum absolute discrepancy of 127,874. The guarded derivation therefore does not fill missing Scottish populations by subtraction. Replaying the identity check matches the cached receipt; this checks the held population table, not a fresh extraction of all WHO mortality files. Correct Boolean decoding and source agreement do not establish exchangeability, reliable vaccine exposure measurement or causal identification.

S72.12 Workbook coverage and effect-measure labels

The seven retained historical result workbooks have a persistent inventory of 294 sheet names and file hashes. Review status does not determine whether a workbook is checked: reclassifying a sheet as a display, duplicate or resolved item cannot remove it from this inventory. Missing files, changed bytes or sheet names, duplicate identities and newly discovered workbooks without a disposition stop the check. This guards the acquired historical workbook set, not an assumed census of all possible data sources.

The current row reconciliation matches 43,106 displayed rows to current CSV records at 11 significant digits, with all compared rows matching on 199 sheets. Header matching is not a complete content audit. Custom-layout sheets and changed historical values require separate source and interpretation checks; agreement with a derived CSV is not independent confirmation of primary data or causal validity.

One custom table, “Direct Individual,” has ten rows and 70 checked numeric or missing-value fields. Seven rows repeat adjusted Poisson-QMLE NHIS associations using overlapping respondents. Three calibration rows require distinct measures despite the table's generic RR heading: the five-study MMR calculation is a historical mixed RR/OR/HR

summary, Taylor et al. (2014) report an OR, and Andersson et al. (2025) report an HR per milligram of vaccine aluminum for combined neurodevelopmental disorders. The latter is not the autism-specific endpoint; matching by author and the shared point estimate 0.93 would conflate the two. The companion records the source identities, actual measures and outcomes without adding any of these repeated summaries to an independent likelihood.

The complete 28-row NHIS summary is also reconciled to 532 fields in its parent estimate table. The semantic key includes outcome, exposure and specification, retaining both adjusted and unweighted sensitivity rows. A repeated display, a changed row order or the same respondents under another specification does not add an independent sample.

Two further presentation tables, “Numerical Ledger” and “Manuscript Table,” contain 24 displayed rows representing 12 historical results. Their companion checks 132 source values or missing fields, 132 cross-display fields and 20 native-result fields. The local MMR displays use fixed-effect intervals for the native coefficients but HKSJ intervals for their illustrative inversions; these conventions are retained explicitly rather than treated as interchangeable. The ADDM coefficient is per six-site uptake SD (2.35 percentage points), not the statewide SD (8.91 points) used in an earlier calibration. Withdrawn PCV conversions and their power claims are not usable individual-risk bounds, including where a source file retains the old numbers for inspection. The interaction-product coefficient is not by itself the total effect of jointly increasing both coverages. These checks do not add new independent likelihoods or validate the causal assumptions of the historical models.

S72.13 Missing values and historical dashboard formulas

Three archived register dashboards contain fifteen displayed result rows representing five repeated families. Their audit traces105 linked numeric, missing and interpretation fields,57 numerical comparisons with held source tables and37 inventory/status totals. A count of historical PASS labels is a count of saved labels, not newly executed validation or evidence of causal identification. The original workbook snapshots remain unchanged.

The21August workbook's literature benchmark displays p=0 because a bare formula references an empty source cell. Its p-value is unavailable. The corrected companion preserves that missingness; no p-value is inferred from the reported ratio and interval. The generator explicitly tests for an empty source cell while preserving genuine numeric zero and propagating source errors. Eighteen in-memory formula-engine and identity checks exercise these distinctions, row reordering and missing/duplicate identifiers. Display text is not an analysis identifier: a replacement childhood-MMR cohort family cannot silently substitute for the historical mixed RR/OR/HR benchmark. These are display-handling corrections, not altered study estimates or additional evidence.

The main historical summary's268 single-SUMIFS formulas each select exactly one source record, whose finite numeric token agrees with the cached result. The screen finds no missing match, summed duplicate, truncated matching row or unexplained numerical difference in that subset. All268 referenced numeric values are stored as text in the workbook; explicit parsing is used for this source comparison. This does not certify recalculation in native Excel. The extended formula and literal-value checks are described in S72.14. No headline likelihood, sample or effect estimate changes as a result of these checks.

S72.14 Summary arithmetic, literal tables and dictionary provenance

All444 formulas in the historical main summary agree with independent source arithmetic. These comprise268 single-record lookups and176 additional counts, conditional aggregates, linked values and arithmetic expressions. Formula dependencies are evaluated from their precedents rather than accepted from saved summary results. Numeric strings and Boolean flags are decoded explicitly. Tests distinguish a legitimate zero from a missing operand or an empty aggregate selection, reject misaligned ranges and unsupported syntax, and expose circular references. These checks establish agreement with the held source tables, not native Excel behavior or fresh primary-source extraction. The original workbooks remain unchanged.

The literal interaction and schedule-revision tables require separate checks because their displayed numbers are text, not formula results or numeric cells. Forty-two display-cell comparisons reproduce the reported point estimates, intervals, p-values, robustness summaries and inventory counts from uniquely identified source rows at the displayed precision. Positive, inverse and nonsignificant results are all retained. The event inventory contains693 records:310 same-year overlaps,188 adjacent-year changes,186 isolated changes and9 without an exact year. The historical summary omitted the last category from its breakdown; the companion supplies it. These are retained inventory records, not missing participants or nine null findings.

The historical data dictionary contains634 field entries. All4,438 associated file, key, field, type, nullability, description and example values reproduce from its JSON source. Identifiers retain leading zeros, and text Boolean tokens are not accepted as Boolean nullability flags. This verifies faithful reproduction of that historical schema, not completeness of the current project's dictionary or correctness of every historical description. None of these repeated displays adds an independent likelihood. Scientific interpretations, remaining custom views and native-application numeric-text recalculation retain separate audit requirements.

S72.15 Power summaries, acquisition counts and workbook guides

Two historical power and acquisition summaries reproduce 16 formulas and 122 source comparisons. The power table's five identification-passing labels are dated screening decisions. Their minimum detectable ratios refer to native model contrasts, including correlated PCV rollout contrasts, not five independent experiments or five individual-risk exclusion bounds. The displayed counts of designs meeting 80% power thresholds of 5%, 10%, 20%, 50% and 100% are respectively 1, 2, 3, 4 and 4 under those historical definitions. These counts do not supersede the exposure-aligned power analyses or current identification audit.

The original frontier inventory's 799 unique payload hashes are reproduced by the pre-restoration base's 720 hashes and 80 retained Swedish acquisition records, which contribute 79 additional hashes and one duplicate. The retained Swedish files were separately hash-checked and restored as described in S72.16. The power summary's total of 805 files and 1,286,180,787 bytes cannot be reproduced from the available frontier and NCES snapshots, whose union has 809 files and 1,286,245,671 bytes. The four-file, 64,884-byte difference remains unresolved. Historical file-validity labels are not fresh proof of parseability, and an unexplained inventory count is not evidence of missing participants or null effects.

Five historical guide sheets receive explicit dispositions for all 187 nonempty rows, with 148 navigation, count and selected numerical claim checks. Three register guides incorrectly describe the Northern Ireland bootstrap as having three report-year clusters. The source reports thirteen report-year clusters for three vaccine additions. The corrected companion and generator retain that distinction without changing coefficients, intervals or bootstrap draws. The generator also uses “selected age,” not language implying preregistration. Dated guide claims, saved PASS labels and repeated effect summaries do not constitute current causal certification or additional observations.

S72.16 Acquisition-record preservation and format checks

The subnational acquisition inventory contains 837 records after restoring 80 reviewed Swedish records omitted by the base-ledger rebuild. Every field of the 757 pre-existing records is preserved in a checksum-pinned snapshot. The restored records contribute 79 unique payload hashes, bringing the inventory to 799; a repeated export is counted once. Re-running the restoration adds no records. Rebuilds include the reviewed extension, while conflicting hashes, missing files, changed provenance and duplicate path identities require review rather than silent substitution.

The restoration checks distinguish byte identity and file format from scientific validity. Four FHM coverage CSVs decode with the source-specific Windows-1252 encoding, preserving Swedish labels. Fifty-six JSON payloads decode, eight XLSX containers pass structural and CRC checks, and twelve older XLS files have the expected OLE signature. The last check does not establish that every worksheet or cell parses correctly. Metadata files and repeated exports are not additional outcome observations. Injected failure tests reject HTML/error payloads, malformed tables, missing files, altered hashes and ambiguous acquisition flags. An existing file attached to a failed retrieval request is not counted as a successful acquisition.

A separate census retains 258 later files absent from this legacy acquisition table, including 129 retrieval sidecars. They span Swedish injury and external-cause data, population denominators and historical vaccination reports. Their absence from this older table does not establish omission from the current analyses. The census preserves paths, hashes and roles for comparison with current source and analysis indexes, without automatically adding likelihood weight. The restoration itself changes no outcome data, fitted effect or synthesis weight.

S72.17 Swedish source coverage and omitted-row sensitivity

The file-level crosswalk accounts for all 258 later Swedish files. The injury analysis consumes 114 outcome series; the pairwise analysis uses three population or county-coverage inputs; one national report supplies manually coded hepatitis-B anchors; nine files contain metadata or query specifications; two national reports provide context; and 129 files record retrieval provenance. These are not 258 new datasets. All 114 API responses contain a complete single page. Their values agree with independent decoding, and no duplicate cell keys or suppression tokens occur. Parser guards reject incomplete pagination, duplicate keys, mismatched dimensions and invalid numeric values. The 46 existing Swedish fitted blocks retain identical numerical inputs. The pairwise design has all 672 required PCV, 147 rotavirus and 210 hepatitis-B exposure keys, with no missing or nonfinite values. Missing required coverage now stops computation rather than becoming zero. Population weights reproduce the denominators; the use of 2024 population for 2025 remains an explicit approximation.

Socialstyrelsen counts a patient once per year, diagnosis group and geographic area. Consequently, sums across diagnosis groups are generally not unique-patient counts. The fracture, burns, poisoning and land-transport composites are summed diagnosis indices. Subtracting T78 and T80–T88 from the chapter total is a signed lower-bound index, not an exact count of patients with other injuries. This distinction is stated in the original analysis notes and confirmed by the publisher's measure definition. The API documentation specifies pagination but does not establish that an omitted record denotes zero. The observed presence of counts of one and two argues against blanket small-count suppression, but cannot resolve every absent record.

A post hoc sensitivity treats absent components as unavailable and applies the same complete-positive-series rule. Six composite samples change. Fractures and the two land-transport composites retain no complete series under this rule and are unavailable, not null estimates. The three remaining composites, matched autism outcomes and autism-minus-composite log outcomes are fitted with the original six covariance/trend variants and three prior specifications. The table reports the reference-prior, model-averaged comparison. Ratios concern the stated diagnosis index; they are not unique-patient injury RRs.

Diagnosis index	Original retained cells	Missing-component sensitivity cells	Original ratio (95% credible interval)	Sensitivity ratio (95% credible interval)
All-injury signed index	756	738	1.056 (0.980–1.132)	1.058 (0.976–1.133)
Burns diagnosis sum	612	72	1.161 (0.916–1.443)	0.881 (0.591–1.225)
Poisoning diagnosis sum	684	288	1.376 (0.930–2.014)	1.414 (0.913–2.080)

The corresponding matched-sample autism estimates are 0.999 (0.774–1.506), 1.034 (0.753–1.453) and 1.154 (0.844–1.596). The differential estimates are 0.959 (0.731–1.571), 1.148 (0.649–2.130) and 0.820 (0.472–1.451). Sample selection, not a change in observed values, produces these differences. These nine comparisons and 108 kernel/prior summaries reuse existing outcomes and receive no additional independent likelihood weight. They do not establish a causal exclusion restriction for injury outcomes. The main autism synthesis is unchanged; using these composites to calibrate design bias requires the omission convention, aggregation definition and sample sensitivity to be addressed explicitly.

S72.18 Missing-component sensitivity of control calibration

Six Swedish rows in the control-calibration model depend on the composite missingness convention: the all-injury, fracture, burns and poisoning indices, and two land-transport indices distinguished by care setting. Under complete-component selection, three control estimates are replaced and three are unavailable. The overlap-filtered specification already omits the all-injury and fracture indices, but retains burns and poisoning; it therefore does not remove this sensitivity. These transformations are based on source identities and sample availability, not the direction or significance of an estimate.

The bias model is refitted under four input policies: the original zero convention, complete-component selection, omission of all six affected rows, and omission of all Swedish control rows. Each policy is evaluated in the primary, overlap-filtered primary, primary-plus-secondary and non-injury-primary sets, with within-context sampling correlations of 0, 0.5 and 0.9. The 48 models retain the original priors and Gaussian reconstruction of control likelihoods. Correlations are assumed, not estimated. A direct multivariate-normal calculation checks the covariance likelihood, and the unchanged baseline reproduces the original bias summary to numerical precision. The non-injury set contains no Swedish observations and remains identical across the four policies.

Primary-control policy	Retained controls	Shared bias mean	Shared bias posterior SD	Sweden predictive bias mean	Sweden predictive bias SD
Original zero convention	45	0.0382	0.0911	0.0560	0.0435
Complete components	44	0.0381	0.0912	0.0545	0.0535
Affected composites omitted	41	0.0383	0.0915	0.0540	0.0651
Swedish controls omitted	29	0.0346	0.1173	0.0346	0.3455

All bias quantities are on the log scale. The table assumes zero sampling correlation; it does not report vaccine-risk estimates. Under correlation 0.9, the Swedish predictive SD is 0.3969 with the original inputs and 0.4092 under complete-component selection. Thus the missing-component convention has modest effects on these mean-bias summaries, while calibration precision depends substantially on the information attributed to related control outcomes. These calculations neither verify that injury outcomes are unaffected by vaccination nor establish that their bias transfers to autism. They concern the control model for the restricted 15-context synthesis, not the expanded 17-context heterogeneous headline.

S72.19 Geographic coverage of Swedish hepatitis-B anchors

The county reports supply 42 hepatitis-B coverage fractions for the 2005 and 2006 birth cohorts. Comparison with the rendered tables confirms the extracted values: the 2005 measure sums the separate three- and four-injection columns, while the 2006 measure uses the combined three-to-four-injection column. Neither is the any-dose percentage. Table 3 of the 2015 annual vaccination report, page 26, supplies five summary percentages, 29.7%, 30.4%, 34.8%, 41.7% and 52.9%, assigned to the 2008–2012 birth cohorts. The final value describes reporting counties excluding Stockholm. The report records 86,338 children and 74.2% population coverage, with Stockholm missing because of technical problems.

The exploratory pairwise reconstruction therefore matches the 2012 anchor to the reporting counties, leaving Stockholm's separately specified path unchanged by that anchor. Relative to imposing the percentage on all counties, this changes 84 county-cohort entries in the interpolated path and 60 in the partially anchored path, confined to the 2012–2015 cohorts. The largest change is 15.3 percentage points. The two corrected paths reproduce 52.9% over the reporting counties using the model's population weights. Their all-county means are 57.5% and 50.4%, respectively; these are reconstructed scenarios, not new nationwide observations. Population rather than actual respondent weights and unobserved county-year interpolation remain limitations. Outcomes, population components, PCV exposure and rotavirus exposure are unchanged.

This reconstruction is distinct from the exposure inputs used by the main joint-country synthesis. All 6,732 cells in its 1,122-row, six-scenario hepatitis-B matrix reproduce from the stated eligibility rules and assumed 95% uptake, without the summary-table anchors. The pooled interaction analysis likewise excludes the Swedish exploratory pairwise estimates. Neither headline changes from this geographic correction. The full pairwise and diagnostic refits are reported below; alternative reconstructions do not constitute independent observations.

S72.20 Calendar-period proxies and Swedish interaction estimates

The 3 September 2012 coverage article attributes county percentages to GSK but describes vaccination during calendar half-years. It does not identify the recipients' birth cohorts, dose counts or the underlying denominator. All nine percentages used in the exploratory reconstruction match the article; assigning them to birth cohorts nevertheless requires an additional assumption. The article assigns 89% to Jönköping and 91% to Örebro. Its underlying GSK dataset is unavailable in the acquired sources. We therefore compare the period-proxy reconstruction with a path that omits those proxies and interpolates between the separately sourced cohort anchors, retaining the reporting-geography calibration. The latter remains an imputed path, not observed county-cohort uptake.

Each reconstruction uses the same 1,122 outcome cells, 5,610 age-weight components, 672 county-cohort units, estimator and random-number sequence. The battery contains 24 pairwise fits, six rotavirus recodings, two county-path permutation diagnostics with 150 draws each and 200 gamma–Poisson recovery simulations. Removing the period proxies changes 189 HepB unit values. PCV and rotavirus inputs and all outcome and denominator values remain fixed.

HepB reconstruction	PCV × HepB multiplier	HepB × rotavirus multiplier
Period proxies; all-county summary calibration	0.446 (0.048–4.134)	0.301 (0.018–4.958)

HepB reconstruction	PCV × HepB multiplier	HepB × rotavirus multiplier
Period proxies; reporting-county calibration	0.470 (0.054–4.068)	0.311 (0.010–9.680)
No period proxies; reporting-county calibration	1.279 (0.216–7.596)	0.296 (0.038–2.317)

The multiplier is the fitted departure from multiplicative separate effects under the specified joint-exposure rule, not the total effect of adding a vaccine to a schedule. PCV–HepB uses the upper joint-exposure rule and HepB–rotavirus the lower rule. Intervals use the county-path permutation SD and a t(20) critical value. With proxies and corrected geography, 138/150 and 32/150 permutation fits are retained; without proxies, the counts are 135/150 and 46/150. Such fit failures and exposure sensitivity limit interpretation. The paired permutation rank probabilities without proxies are 0.743 and 0.191. They are diagnostics, not proof of exchangeability or absence of an effect.

The PCV–rotavirus margin is unchanged by this HepB sensitivity. Two alternative rotavirus recodings yield HepB–rotavirus multipliers of 0.230 (0.062–0.848) and 0.212 (0.059–0.765) without the period proxies. These are retained despite excluding one under their shifted-exposure/cluster interval calculation; they remain weakly identified, imputation-dependent recodings and do not establish protection. Under simulated zero interaction, 199/200 recovery fits are retained: the mean log multiplier is 0.104, its simulation SD is 0.495 and nominal 95% cluster-interval coverage is 91.0%. With proxies and corrected geography, the corresponding values are 0.094, 0.504 and 93.5%. These are simulation diagnostics rather than additional empirical effects. None of these overlapping rows adds independent weight to the pooled interaction or autism synthesis.

The Örebro 2015 child-health report, pages 10 and 42–43, supplies a distinct primary-source check. It reports HepB coverage of 97.3% and PCV coverage of 96.3% for births in 2015, MMR coverage of 97.2% for births in 2013 and diphtheria coverage of 97.1% for births in 2014. The vaccination table counts resident children with one to three injections; it is not a completed-course table. Eligibility for free HepB vaccination of all newborns begins with births on 1 January 2013. The report's national chart comparator is births in 2012, not the county's 2015 cohort. The BHVsystem records feed the national agencies, precluding treatment of county and national summaries as independent samples. These records are retained with their cohort, dose, denominator and overlap definitions rather than inserted as interchangeable exposure anchors.

S72.21 Numerical limits of the complete-component calibration sensitivity

The complete-component calibration uses 44 controls in the restricted 15-context common-effect model. Increasing the number of importance-sampling draws fourfold and varying the bias-distribution quadrature gives the following numerical sensitivities. Each row uses two independently scrambled sequences; the entries are repeated calculations of the same model, not independent evidence.

Draws per scramble	Bias-distribution nodes	Model ratio (95% interval)	Minimum relevant core ESS
262,144	75	0.873 (0.741–1.025)	8.26
262,144	420	0.872 (0.741–1.024)	3.59
1,048,576	420	0.871 (0.741–1.022)	16.37

All three calculations remain below the numerical ESS floor of 50. The higher-draw calculation's two scramble-specific upper limits are 1.024 and 1.019; their log difference is 0.00485. Stability of these summaries and exact reintegration of eighteen residual-shared-bias scenarios establish reproduction of the saved calculations, not sufficient importance-sampling precision. The coarse/fine comparison is a quadrature sensitivity check, not proof of convergence at finer resolutions. These provisional intervals cannot support a tight exclusion bound. More efficient importance sampling and further quadrature checks are required before substantive interpretation. Control-outcome exchangeability remains a separate identifying assumption even if numerical precision is improved. None of these calculations replaces the 17-context heterogeneous estimate or adds independent evidence.

S72.22 Municipality identities and LSS source checks

Swedish municipality identifiers retain diacritics: Habo (0643) and Håbo (0305) are distinct places. Numeric identifiers may also carry footnotes, including Sorsele's `2422**`. The parser checks agreement between codes and names, rejects duplicate or conflicting matches, and preserves suppression as missing rather than zero. An independent worksheet reader compares 60,610 fields across the nineteen 2007–2025 annual workbooks. The resulting panel has 5,510 municipality-year rows. Combined person-circle totals are available only when all required components are observed.

The municipal battery includes fixed effects, differential trends, event studies, county-assignment diagnostics, synthetic-control paths and coverage associations. Selected estimates below concern recipients per 10,000, with county-cluster standard errors and small-cluster t intervals. They are not individual vaccine risks.

Rollout and outcome	Municipality/year fixed effects	With differential linear trend
Early rotavirus; recipients aged 0–22	10.589 (6.565–14.614)	3.506 (-3.082–10.094)
Early rotavirus; person circles 1+2	1.066 (-0.563–2.695)	0.817 (-1.582–3.216)
Norrbotten HepB; recipients aged 0–22	-5.533 (-9.832–-1.235)	6.251 (3.284–9.219)

All eight joint pretrend tests reject at 5%, across the two rollouts and four service-use outcomes. Changes in specification can change sign as well as magnitude. Positive and negative associations are retained, but neither the rejection pattern nor the source checks establish an identified vaccine effect. LSS measures service use; person circle 1 includes several diagnoses, and other person circles are not automatically valid negative controls. Historical power and proportional-stock conversions remain conditional diagnostics rather than validated individual-risk exclusion bounds.

The separate county likelihood and SC/SDID/ASCM analyses use 205 municipalities with complete usable count/rate records throughout 2007–2025. The affected identity and footnote cases remain outside this sample because other years are unavailable. All 3,895 underlying municipality-year outcome cells agree under the two source reconstructions; the full synthetic-control battery reproduces 1,554 numeric quantities exactly. Forty-eight likelihood and future-exposure summaries differ only by floating-point roundoff. The distinct autism age-band analysis's 2008–2015 county weights are also unchanged in all 21 counties. These checks add no independent likelihoods. Complete-record selection, rounded-rate implied denominators, municipality-of-provision counts, donor exposure and diagnostic ascertainment remain separate limitations.

Header-based extraction verifies 76 national age totals and independently reconstructs all 437 annual single-age weights. Uniform allocation within the published age bands is an assumption, not observed age-specific vaccination among LSS recipients. The Kolada N28891 definition identifies Socialstyrelsen as its source, so agreement between Kolada and the annual workbooks is a dissemination check rather than independent outcome replication. Among 3,722 overlapping municipality-year rates, 91.0% agree within one recipient per 10,000; some discrepancies exceed rounding, with a maximum of 85.59. Another 295 Kolada rates have no usable workbook counterpart. Two of these concern municipality-years that the publisher identifies as lacking valid reports. The remaining 293 are retained as additional same-register rates, not substituted for suppressed counts or added as independent evidence. The reporting-quality and alternative-panel checks follow below.

S72.23 Reporting quality and alternative dissemination of Swedish LSS rates

Numeric cells must be interpreted alongside the publisher's reporting notes. The 2022 age worksheet records zeros for Tranås, but its footer identifies that municipality as nonreporting. Its count and rate are therefore unavailable, not observations of zero service use. The source-cell ledger preserves the printed zeros and the footer separately from the analysis values. Genuine zeros elsewhere remain zero. The 2022 quality declaration, section 2.2.4, independently confirms that Tranås did not supply valid register information.

Footnote scope also matters. The 2020 quality declaration reports complete municipal response. The 2021 declaration identifies Söderköping as the exception. Sorsele's nonresponse note in the adjacent historical national table is not a basis for excluding its current 2020–2021 observations. The 2015 workbook separately states that Lidingö did not report and that its previous-year values were used in Stockholm county and national totals; its municipal rows remain missing. National age totals can therefore contain publisher imputation and are not certified as complete contemporaneous measurement merely because the cells reproduce.

Fresh downloads reproduce the held 2018, 2022 and 2024 workbook bytes and their selected Kolada API rates. Nordmaling's 2018 workbook rate is 12 per 10,000 versus 97.589 in Kolada; Lysekil's 2024 rates are 24 and 61.038. These differences are not stale-cache artifacts, and their cause is unresolved. Kolada supplies rates of 104.849 for Söderköping in 2021 and 79.559 for Tranås in 2022 despite the official nonresponse statements. Both are retained in the source ledger but excluded from the alternative analysis panel. No counts are inferred from these rates.

The alternative panel retains 4,015 municipal rates for 2012–2025. Four versions distinguish dissemination and selection: all usable workbook rates, all usable Kolada rates, and each source restricted to the same 3,722 municipality-year keys. The analysis reports 120 overlapping specifications, 240 event-study rows including reference years, and 20 exhaustive county-assignment diagnostics. Five windows cross the 2014 partial-rollout or 2015 cohort cutoff with follow-up ending in 2015, 2018 or 2019. Each uses municipality/year fixed effects, a differential-linear-trend alternative and a complete-municipality sensitivity, on native rates and log rates. Later years contribute to 168 annual descriptive summaries, not a comparison against allegedly unexposed counties after national rollout.

Rate source, 2012–2015; 2014 cutoff	Fixed effects	Differential linear trend
Annual workbook	7.041 (4.733–9.348)	-0.133 (-9.755–9.490)
Kolada	7.043 (4.736–9.350)	-0.125 (-9.730–9.480)

Entries are differences in LSS recipients aged 0–22 per 10,000, with county-cluster t intervals. The two sources agree closely in this window, but the estimated difference is sensitive to trend specification. The native-rate pretrend comparison has p=0.135; the corresponding log-rate comparison has p=0.0034. With only two pre-rollout years, nonrejection on the native scale offers little assurance. The plug-in 80%-power minimum detectable difference is approximately 3.26 per 10,000 for fixed effects and 13.59 with the differential trend in the workbook panel. These are conditional noncentral-t calculations using fitted standard errors, not individual-risk power or validated small-treated-cluster operating characteristics.

Across the full 120-row sensitivity set, 79 nominal intervals lie above zero, 12 below zero and 29 include zero. These are correlated specifications with common source data, not counts of independent replications or known false positives. Positive, inverse and null-compatible results are all retained. The fit windows use 2,303 Kolada municipality-year rates; the later retained rates contribute descriptive information but do not create an untreated post-national-rollout comparison.

Stockholm and Jönköping are the only treated counties. The outcome mixes many birth cohorts unaffected by infant rollout. A 2015 cutoff places partial 2014 exposure in the preperiod; later donor adoption and the 2019 national programme complicate longer windows. County assignments were not randomized, so the assignment distribution is diagnostic rather than a causal randomization test. Kolada starts in 2012 and cannot reproduce Norrbotten's 2012 HepB introduction comparison with a pre-introduction period. These explicit limitations preclude treating the additional rows as independent causal effects or individual vaccine risks. They add no independent weight to the autism synthesis.

S72.24 Geographic aggregates and organizational-unit availability

The retained Kolada N28891 exports contain 53,952 usable total-rate records: 4,017 municipality-year records, 45,864 records for groups with current membership metadata, 4,060 group records without current membership metadata, and eleven national totals. The group records are not additional municipalities. All are retained with explicit roles, but none is added to the municipal panel as an independent population.

The version 3 API specification identifies G-prefixed codes as municipality groups. The complete current metadata response contains 3,343 groups; 3,276 occur in the held N28891 exports. Direct ID queries also fail to recover membership for the remaining 290 groups. The publisher's API documentation distinguishes municipality-level from organizational-unit data. N28891's retrieved `has_ou_data` flag is false: this particular indicator does not advertise a school-, facility- or other organizational-unit series. That does not establish the absence of organizational-unit data for other Kolada indicators.

A dissemination check reconstructs unweighted means from the currently listed municipal members and held municipal rates, including genuine zeros. It reproduces 23,647 of the 45,864 documented-group values within 0.000001 per 10,000; contributor counts agree in 44,582 records. These are accounting diagnostics, not a certification of historical group membership. Current memberships and older disseminated aggregates can differ; unresolved discrepancies remain visible and are not repaired by selecting whichever members reproduce a desired mean. The API contributor-count field is not a count of patients. Group aggregates may carry information unavailable in individual municipal releases, but

exploiting that information would require verified membership, aggregation rules and the dependence structure. Treating these overlapping aggregates as independent geographic observations would overstate precision.

S72.25 Pennsylvania source layouts and complete county coverage

The twelve acquired Pennsylvania county school-immunization workbooks cover 2014–2025. Some county blocks place an entire multiline header in one cell; some also place a complete kindergarten count or percentage row in one cell. The reader validates the vaccine order within those blocks and checks each published rate against its vaccination numerator and enrollment denominator. A blank percentage remains unavailable; a numeric zero remains zero. An unrecognized schema or ambiguous row alignment is an error, not a reason to silently skip a county. The 2020–2021 county report contains examples of both layouts.

The panel contains 2,412 rates: 67 counties, twelve school years, and three vaccine series (MMR, hepatitis B and two-dose varicella). Fifty-seven rates occur in nineteen text-collapsed county-year blocks. All 2,412 rates agree with the source counts to the source rounding precision; the largest absolute discrepancy is 0.000499 on the proportion scale. The separately reported “Out of State” block is retained in the audit but is not a Pennsylvania county. These checks concern the complete county inventory in the acquired reports, not complete ascertainment of every child or every possible historical report.

County analysis	Ratio	95% interval	County-years
Same-year MMR, county and year fixed effects	0.990	0.887–1.106	670
Same-year MMR, county-specific linear trends	1.019	0.870–1.194	670
MMR×HepB interaction term	0.919	0.654–1.290	670
Federal-outcome MMR×HepB interaction term	1.114	0.808–1.535	335

The first two rows describe autism's share of special-education enrollment per ten-percentage-point difference in county kindergarten MMR coverage. The interaction row exponentiates the product coefficient for within-county-centered MMR and HepB coverage, each measured in ten-percentage-point units. It is not the risk ratio for individual receipt of both vaccines. All intervals use county-cluster uncertainty with a t critical value based on 67 counties. Other disability outcomes are analyzed as alternative outcomes, not automatically classified as negative controls.

The source-complete battery contains 444 PennData specifications, 36 federal-outcome specifications and six interaction-model terms across the two reporting systems. The state and federal series cover overlapping populations, so they are alternative reporting systems rather than independent cohorts. Federal joins contain 1,005 matched vaccine-county-year rows (335 county-years for each of three vaccines); exposure lags refer to exact calendar years. The federal interaction's multiple-testing adjustment retains the complete original 15-term family. Kindergarten exposure is a cohort flow, whereas the outcome is a broad school-age stock. Source completeness does not resolve this mismatch or establish exogenous uptake.

In the all-jurisdiction kindergarten-MMR composition family, the reference working-bias model gives 0.998476 (0.945856–1.053979). The 49 local-family estimator and assumption combinations preserve the original jurisdiction counts and source-overlap rules. These local gradients and their repeated specifications add no independent weight to the distinct 17-context autism synthesis or the 27-pair interaction synthesis. The source-cell ledger, acquisition routes, paired model replays and repeated-display crosswalk are supplied by analyses 3496–3499.

The independently constructed age-specific analysis in S45.5 uses the same complete 804 county-year MMR count pairs. Its vaccination and enrollment counts agree exactly with the county extraction. The age-specific model uses enrollment-weighted counts and separately printed disability denominators; the county model uses source-published coverage and county outcomes. These are distinct, overlapping analyses. Eighteen accompanying power scenarios retain their explicit assumption of thirteen equally represented cohorts; the source model also supplies separate seventeen-cohort scenarios. Neither assumption is a verified mapping from kindergarten coverage to individual exposure throughout the observed school-age stock. These power calculations therefore do not establish individual-risk exclusion bounds.

S72.26. Acquisition counts and duplicate source URLs

Acquisition records and unique data payloads are counted separately. Distinct URLs can refer to a single retained file, and multiple tables can describe the same population. The acquisition merge retains source URLs and assigns one counted representative per eligible content hash. Conflicting byte counts, conflicting file versions at one path, missing hashes and ambiguous status flags cause the merge to fail.

The historical frontier inventory and NCES acquisition records contain 809 distinct eligible hashes totaling 1,286,245,671 bytes. A prior 805-file total is exactly reproduced by retaining the last hash/path duplicate without carrying forward its counted status. Four New Jersey statewide discipline workbooks, totaling 64,884 bytes, account for the entire difference. Each has two recorded source URLs, one counted record and one uncounted alias. All four payloads remain in the current acquisition inventory, and their bytes match the historical hashes. The ten NCES geographic-linkage archives also match their acquisition hashes. This count describes the specified historical inventory union, not all data in the project.

Independent retrieval of all eight recorded New Jersey URLs distinguishes the historical inventory arithmetic from source identity. The URLs return eight distinct workbooks: four reporting margins for 2021–22 and four for 2022–23. A basename-only cache had incorrectly attributed the later-year payload to each earlier-year URL because the directories reuse the same filenames. The source register retains the historical attribution alongside the independently retrieved URL, payload hash and cell-level extraction. URL-specific filenames prevent recurrence in the New Jersey acquisition function. This check covers the eight specified URLs, not every acquisition routine in the project.

The New Jersey files tabulate disciplinary removals among IDEA-served children aged 3–21 by disability, sex, English-learner status and race. Analysis 3501 extracts 1,456 measurement cells, including 192 suppressed cells retained as unavailable and 476 observed zeros. The four margins agree on statewide total removals of 29,543 in 2021–22 and 33,815 in 2022–23. The disability-table autism rows report 748 and 932 removals, respectively; these are neither new autism diagnoses nor counts of distinct children. A printed 2021–2023 title in the later English-learner workbook is retained as inconsistent with its filename and companion tables. The four margins overlap and are not independent cohorts. These tables alone supply neither cohort-aligned vaccination exposure, a matched population denominator nor an adequate longitudinal identifying comparison. They are retained as descriptive evidence with zero additional vaccine-effect likelihood weight. Analyses 3500–3501 provide the reconciliation, acquisition receipt, source hashes, extraction and regression tests. The acquisition correction changes no fitted vaccine-effect estimate.

S72.27. URL identity and saved source pages

Analysis 3502 examines the headers of 2,231 top-level CSV files and traces explicit URL/path fields in 150 tables containing 696,638 records. Repeated observations are collapsed only for the provenance census, retaining their table, field names, first and last row and reference count. The resulting 9,190 relationships identify 3,474 recorded paths. Twelve paths have multiple URLs: the four New Jersey discipline collisions, seven catalog-versus-direct-download references, and a shared Virginia HTML filename. Textual locators, missing paths and multiple paths in one cell receive separate dispositions. Thirty-seven tables with other URL/file schemas receive the source-specific review in S72.28. Nested manifests and unrecorded downloads are outside this check.

The Virginia acquisition log records four different page hashes at one path. The acquisition code downloads and parses each page before requesting the next, so the overwrite establishes lost source-page snapshots, not loss of the links already processed. URL-specific source-index filenames prevent that collision. Four current VDH pages are retained separately; none is represented as a recovery of its historical bytes. Three archived Tableau workbook identifiers recovered from those pages return HTTP 404 at their workbook endpoints. This is a recorded retrieval limitation, not evidence that alternative archives or data are unavailable. The retained ImmunizeVA extracts cite the same VDH school survey and are not independent collections.

The current VDH release has distinct measurement rules: coverage is published in bands, small-enrollment values are suppressed, and nonresponse is labeled separately. The VDH release notice identifies 2025–26 as the first school year collecting kindergarten vaccine-group rates in addition to overall required-schedule completion. Historical completion percentages cannot therefore be relabeled antigen-specific kindergarten coverage. Current bands cannot be imported as exact rates, and neither suppressed nor unreported values can be replaced with zero. This source review supplies no new identified outcome effect or independent likelihood weight; extraction and harmonization of the current release remain separate work.

S72.28. Source fields, parent records and search coverage

The current subnational acquisition view retains 837 source records and 803 eligible payload hashes, with URL-specific attribution for all eight New Jersey discipline workbooks. Historical source identities remain available beside current identities. This count concerns the specified subnational inventory, not the full project or 803 independent datasets. Rebuilding the view preserves all source records and reproduces the same URL-to-payload mapping.

The 37 additional schemas contain 163,224 records. Explicit parent keys distinguish the exposure and outcome files in 86,856 Dutch joined rows and the vaccine-source URL from the outcome-source URL in 148 Nigerian rows. Five Costa Rican tables link 2,389 report/page references to their document identifiers and URLs. Keyword matches are discovery candidates rather than validated vaccine measurements.

Ohio's 26,918 building-outcome rows and 12,195 matched exposure rows trace to their annual acquisition records; 3,348 unmatched exposure rows have unavailable outcomes. Fifty UAE resource records trace to thirteen held dataset-detail pages rather than being represented as successful resource-file downloads.

The resulting 2,338 source relationships refer to 2,120 held files. All 71 distinct table/reference pairs in the provenance and analysis-pointer checks resolve to held documents. File existence and content-hash agreement establish the checked provenance, not the correctness of every source measurement. Twelve profile-metadata schemas, eight readiness-flag schemas and the mechanism-pathway schema do not contain local filenames despite lexical matches to “file” or “path.” Analysis and provenance references retain their distinct roles.

Search coverage is reported separately. The Bing audit has 1,984 cached responses; the Yahoo audit has six cached responses and 1,978 entries without a cached response. Neither a cached search miss nor an unattempted search establishes that no data exist. Search caches are not outcome datasets and receive no likelihood weight. These source-field checks do not certify nested acquisition metadata, all historical source transcriptions or worldwide evidence exhaustion.

S72.29. Unknown enrollment in school-level exposure aggregation

Ohio's school exposure aggregation retains missing enrollment and coverage components as unavailable. Enrollment-weighted coverage requires all constituent school records to have observed weights and coverage. An observed zero percentage remains zero; an unknown enrollment is not replaced by zero, and a zero denominator does not produce an estimated rate. Missing or unparseable exposure identifiers cause the calculation to fail rather than silently dropping records. Identified schools without an outcome match remain in the join with unavailable outcomes.

The Ohio Virtual Academy records for 2020/21 and 2021/22 have reported completion percentages but unknown kindergarten enrollment. Their aggregate enrollment and weighted-completion amounts are therefore unavailable. The source-preserving comparison identifies four affected weight cells across these two school-years. All 15,543 joined row identities, every finite weighted-coverage value and all outcome values agree with the held comparison.

The two affected weighted-coverage values were already unavailable, so this correction adds no model-eligible observation or effect estimate. The outcome remains whole-building disability enrollment rather than autism-specific incidence, and kindergarten coverage is not the exposure history of all enrolled pupils.

S72.30. UAE preview completeness and dose aggregation

The Bayanat source audit enumerates 52 resources on thirteen held dataset-detail pages. Two 2024–2025 metadata resources contain an unescaped apostrophe in their HTML attributes. Reading the complete embedded JSON object recovers five records from each; twelve remaining metadata records on four pagination pages are not held. Absence of a numerical preview is not evidence that a linked dataset cannot be acquired. The source ledger retains all 608 available preview rows and 113 saved pagination responses, with explicit partial-resource dispositions. Column identities are normalized consistently on embedded and subsequent pages, recovering five 2006 student-enrollment records and five 2015 center-count records. The older disability tables contain 91 observations with separate measurement labels. Their years do not overlap the 2019–2022 vaccine tables.

The vaccination ledger retains 828 emirate-level measurements: 564 dose-detail cells, 240 vaccine subtotals and 24 grand totals. Six reported zeros remain zero. Analytical aggregates use dose details only; grand totals are not assigned to the preceding vaccine category. All 240 subtotals agree with their constituent dose counts. The published 2022 MMR dose-2 total-column value differs from its six displayed emirate components by eight doses; both values are retained.

Sums across categories are reported category doses, not necessarily unique injections, vaccinated children or population coverage.

The descriptive first-difference regressions use six emirates and two overlapping years, 2019 and 2020. The outcome is government rehabilitation-center enrollment by the child's reported emirate, not autism incidence. Dubai's exposure covers EHS facilities only and Abu Dhabi has no vaccine-table row. The analysis has no contemporary population denominator or pretrend test and does not identify an introduction or withdrawal effect.

Exposure	Log-change coefficient	Nominal 95% HC1 normal interval	HC1 t(4) sensitivity interval
Sum of reported vaccine categories	0.441	-0.272 to 1.153	-0.569 to 1.450
MMR category doses	0.132	0.018 to 0.246	-0.029 to 0.294
DTP-combination category doses	0.296	-0.488 to 1.080	-0.815 to 1.406

These coefficients measure change in log enrollment per change in log category doses, with a 0.5 count offset; their null value is zero, not one. The normal-reference MMR association is positive (p=0.023); the t(4) sensitivity gives p=0.085. With six units, neither approximation establishes reliable causal coverage. Using category subtotals or the grand total separately reproduces the applicable dose-detail regressions. Counting all levels together multiplies all-category exposures by three and MMR/DTP exposures by two in the overlap years; log differences largely cancel these factors, apart from the count offset. The corrected screens remain descriptive and contribute no additional autism-synthesis likelihood.

S72.31. UAE official-source extensions and access failures

The public download check accounts for every resource in the 52-record Bayanat inventory. Forty-eight records supply distinct download URLs; four have no download link in the held metadata. Forty-four requests redirect to an error page that returns HTTP 200. Their visible page content is identical despite differing response hashes, and none is classified as a data file. Four legacy-host requests fail to connect. These retrieval outcomes do not invalidate the separately held numerical previews and do not establish that the underlying data are unavailable elsewhere. Three candidate EHS vaccination PDFs also remain unacquired locally; cached search text without verified chart labels is not converted into emirate-level observations.

The Ministry of Community Empowerment statistics page supplies embedded annual records for seven emirates. The extraction retains 84 nonprofit-organization records for 2014–2025 and 77 social-welfare records for 2014–2024. English and Arabic numeric values agree for all 161 records. Source labels, geographic IDs, years and units are preserved. Ajman's English nonprofit label includes “Public,” whereas its Arabic category matches the other emirates; equivalent organizational scope is not assumed. These series describe administrative context, not autism or disability diagnoses, and do not provide a diagnosis-risk denominator.

The same page labels its rehabilitation-center statistics 2022–2023. Seven emirate counts sum to 114 centers. Four sector counts sum to 120 in English and 115 in Arabic because the semi-government category is 45 in English and 40 in Arabic. The records retain both language values and both unreconciled margin differences. Center stocks are not student counts. Separately published male and female student shares are 72% and 28%, with no absolute denominator. These composition shares are not prevalence rates; the reporting period is not reassigned to a single calendar year.

The Abu Dhabi Public Health Centre campaign report, dated 10 February 2025, reports more than 89,000 MMR doses administered during 2024 to children aged one to seven. This is a supplementary-dose campaign on an existing routine schedule, whose MMR doses are given at 12 and 18 months. The record preserves the dose-count inequality, administration year, publication date and 2024–2025 school-delivery period separately. It does not equate doses with unique children or universal uptake. No matched diagnosis outcome, eligible-population denominator or cohort-specific uptake contrast is held for this event. It is retained as an exposure candidate, not a new effect estimate or an additional independent likelihood. Analyses 3508–3511 reproduce the extraction and source dispositions from preserved response bytes without submitting forms or executing page scripts.

S72.32. Colorado source vintages, denominators and federal overlap

Five readable Colorado Department of Education administrative-unit workbooks retain 273,656 count and percentage cells for count years 2016–2024. The tables distinguish students reported by an administrative unit from students served by its districts. Reported counts feed the federal special-education collection; served counts exclude some parent-choice private-school placements. Disability categories, grade, sex, race, English-learner status and educational environment are separate marginal tables. Their coexistence does not provide an age-by-autism cross-tabulation. State totals and within-unit totals remain separate from category details.

Every retained cell carries its workbook, sheet, row, column, original token, population scope and measurement role. Missing values and suppression remain unavailable. A suppression marker can reflect a small complementary cell, so it is not interpreted as a simple upper bound on the displayed category. Published percentage inequalities remain inequalities, without midpoint substitution. Nine published percentage values are outside the allowable range. Among percentages checkable against their published counts and totals, 97,555 agree numerically, eight differences are consistent with the source's numerical precision, and 22 disagree. Original values and count-based comparisons are retained separately; none is silently repaired or converted to zero.

Formatting irregularities require explicit reconciliation. Two rows missing administrative-unit codes are matched by unique exact names within the same worksheet, recovering identities for 40 cells. One unlabeled total row is verified against totals in other sheets, accounting for another 20 cells. In the 2020–2024 workbook's educational-environment sheet, two measurement blocks carry the same 2022 headings. All 4,808 affected cells are retained with ambiguous analytical years rather than assigned a year from column position. The other worksheets retain their stated years.

Adjacent-vintage comparisons include 152,348 cell relationships. There are 100,520 exact observed agreements and 3,773 observed differences; the latter occur between the two earliest readable vintages. The remaining relationships include unavailable values, changed availability and unmatched keys, not additional numerical agreements. A separate 2016–2019 archival workbook remains truncated at 1,048,576 bytes. A fresh download reproduces its exact hash, and a byte-range response reports the same total resource size. Overlapping calendar coverage therefore does not establish that the readable workbooks are identical replacements for that vintage.

The state–federal comparison uses 682 federal administrative-unit count records for 2020–2024 and 3,258 comparisons across state vintages and population scopes. Autism age components are combined only when both federal counts are publicly numeric. For the state-reported scope, 1,177 comparisons agree, 34 differ, 410 are not numerically comparable and eight lack a federal unit match. Repeated appearances of the same unit-year are not independent observations. The served-scope comparison has more differences, consistent with its different population definition; it is not substituted for reported counts. The 34 reported-scope discrepancies remain explicit and are not resolved by selecting the value closest to a null effect.

The Colorado model loaders in analyses 127 and 153 use federal IDEA files, not these state workbooks. The workbook percentage and heading defects consequently do not enter those fitted outcomes through the inspected loaders. The state comparison provides a measurement and provenance check, not an independent participant sample, a new vaccine-effect estimate or evidence of general-population autism incidence. Analyses 3512–3513 preserve the original files and reproduce the extraction and comparisons offline.

Fourteen unreadable Rhode Island paths represent six distinct encrypted workbook payloads, identified by file hashes and the state catalogue. Their aliases are retained for provenance but are not counted as separate missing studies. No outcome value or statistical null is assigned to these inaccessible cells. These source checks do not resolve the remaining encrypted-source gap.

S72.33. Missing components in the Colorado other-disability screen

The four-category school-age composite includes hearing impairment, orthopedic impairment, visual impairment and traumatic brain injury. These are additional disability outcomes, not assumed negative controls. The federal Colorado extract has 711 local-education-agency-year records: 378 lack all four components, 251 have some components missing and 82 report all four. A sum of available values cannot be interpreted as the complete composite when components are suppressed or absent. An all-missing sum is unavailable, not zero. Most all-missing rows are ordinary-district records introduced into later files; their presence alone does not mean that each affects a fitted county observation.

The available-component sensitivity preserves an all-missing sum as missing. With the existing exposure windows and county aggregation, it uses 134 county-years in 34 counties and gives a composition ratio of 1.095 per ten-percentage-point difference in kindergarten MMR coverage (95% county-clustered normal-reference confidence interval 0.679–1.766). This remains a sum of reported components relative to special-education enrollment, not the incidence of the four conditions or an individual vaccine-risk ratio. It is not a test that the complete disability composite is unaffected.

Requiring all four components produces 39 eligible county-years in fourteen counties. Both this specification and a version restricting the denominator to units with complete components fail the existing fifteen-county minimum and therefore have no fitted effect estimate. Neither is coded as a null. Analysis 3514 retains the component-level missingness, county panels, original-reference comparisons and unsuccessful-fit dispositions. Matched numerator/denominator coverage and unit-specific exposure-window completeness require separate assessment; these screens add no independent likelihood to the main synthesis.

S73. Descriptive source companions and verification limits

S73.1 Taiwan age-specific autism certificate stocks

National Ministry of Health and Welfare tables and Ministry of the Interior December registered-resident counts provide 95 matched year-by-age rates in a 110-row panel for 2015-2025. The ten age bands run from 0-2 to 65 and older. All ten are available for 2015-2018 and 2021-2025; 2019 supplies five under-18 bands only. The ten 2020 rows contain population denominators but no age-specific autism numerator. Ten 2014 count rows are retained separately without an imputed denominator. Missing adult or year-specific outcomes remain missing, not zero.

The denominator aggregation reconciles the 22 county/city totals to the publisher's national single-age counts, retains the explicit 100-and-older column, and verifies male plus female equals total. The 2019 child count is joined once to the combined-sex denominator. These national and county totals come from one MOI collection, not independent population samples. A separate validator checks all 110 denominators against native national cells, the 95 rates against supplied numerators, and 180 modern workbook count/category cells. Older PDF-cell interpretation relies on its named source audits; arithmetic reproduction is not a full independent retranscription of every historical table.

These rates concern certificate holders at annual snapshots, not incident autism diagnoses, all autistic people, education-service enrollment or county-specific autism outcomes. Since 2013 the source combines old disability manuals and newer identification records. The 2018 yearbook footnote requires existing manual holders to reassess and exchange documents within a seven-year transition following 11 July 2012; it does not establish universal recurring seven-year recertification. Age, classification, eligibility, ascertainment and stock accumulation may therefore change the observed profiles independently of vaccination. No vaccine-effect model is fitted to these 95 descriptive rates, and none enters the autism likelihood.

Primary sources: MOHW national disability tables, MOI household-registration population tables, and the second CRPD report attachment. Exact annual URLs, source pages and hashes accompany `qa/data_handling_20261002/taiwan_age_rate_panel_3619/run_002/receipt.json`, `taiwan_autism_certificate_stock_rates.csv` and the matched `taiwan_rate_validation_3622/run_002/receipt.json`. General-population, Indigenous-population and institutional tables retain distinct scopes; similarly named series are not substituted for one another.

S73.2 Costa Rican local coverage tables

The retained local health-situation reports supply 42 source-bound coverage cells, including 32 outside the historical strict extraction subset. Thirty-five Santa Cruz cells cover 2014-2018; 25 had labels not recognized by the stricter extraction. Four Abangares cells have no established vaccine year. Three Naranjo cells retain header years 2018-2020 and the conflicting caption years 2012-2014. Nearby event prose does not settle a table's year. Six published percentages disagree with the count-implied arithmetic and both versions remain available.

Grecia and Río Cuarto report aliases point to identical bytes; some measurements name Pital and Aguas Zarcas rather than the report locality. The geography is not silently assigned from a filename. Text and table-coordinate inspection covered 139 candidate pages, with selected target tables visually checked; this is not semantic certification of every candidate. No compatible outcome join or vaccine-effect fit is added. The official Santa Cruz and Abangares reports and

exact cell locators are retained in `qa/data_handling_20261002/costa_rica_parser_3624/run_005/receipt.json` and its normalized tables.

S73.3 UAE service counts and Virginia dashboard observations

Ten vaccine-type values in indexed text of official UAE 2024 EHS reports sum to 865,825 services. Six emirate-chart values have the same sum but lack verified label-value pairing and remain unassigned. A secondary account gives 865,852, a difference of 27, and an internally inconsistent annual subtraction. Its 2023 total of 786,187 also differs from the indexed official antigen sum of 779,771. No correction is selected by convenience. The official PDF bytes and chart positions remain unverified locally; unresolved 2021-2023 year-to-bar assignments supply no annual trend rows. These are service/dose counts, not unique children or national vaccination coverage. Direct PDF/XML transport failures do not establish that data cannot be obtained by another public route. Source URLs, indexed captures and discrepancy tables accompany `qa/data_handling_20261002/uae_remaining_3623/summary.json`.

The Virginia 2025 school immunization dashboard displays adequately immunized rates of 90.7% for all kindergarten schools, 91.0% for public kindergarten schools and 90.5% for public seventh-grade schools under their respective filters. These are manual transcriptions of browser text, not independent downloaded-data replay. Grades and school types are not interchangeable. The observed export controls offer images and presentation formats but no data table; the map's View Data control is disabled. The full school table remains unresolved, not proven unavailable through every public route. The collection is the same VDH School Immunization Survey underlying ImmunizeVA, so these observations add no independent replication or effect estimate. Exact selections and export dispositions appear in `qa/data_handling_20261002/virginia_full_export_3626/receipt.json`.

S73.4 Brazilian registered-birth denominators

The 594 SINASC state-year DBC archives comprise 270 files for 1996-2005 and 324 for 2006-2017, covering 27 states in each year. A source-pinned production replay rehashed and decoded every held file and reproduced the saved aggregate parts and both normalization snapshots, totaling 65,873,856 active birth registrations. Invalid dates and unresolved residence groups are retained; annual-only dates are not assigned invented months. The source inventory's metadata-only check has a narrower purpose and does not imply the absence of this all-file replay.

The full replay uses the production decoder. Separate field checks share that decoder, and clean staged dependency testing covers two files rather than all 594. These checks do not establish independent full-scale decoder validation, unique-person follow-up or complete birth registration. The aggregates are supplementary registered-birth denominators for hospitalization and cohort sensitivities: admissions per original registered birth are not autism incidence or current-resident risk. The archives are not 594 studies and add no independent likelihood. The source chain, date-quality groups and remaining limits are documented in `qa/data_handling_20261002/brazil_provenance_scope_3625/REPORT.md`; the all-file replay receipt is `qa/completeness_20260907/age_control_designs/sinasc_reparse_runs/20260908T072444732710/receipt.json`.

S73.5 Reproduction scope

The compact package supplies source extracts, contrasts and saved likelihoods, together with integrity, native-family, substate and diagnostic replay commands. It contains 1,844 manifest-listed files, excluding the manifest itself, and provides separate replay tiers and final integrity verification. The report contains 675 specified checks, including source arithmetic, missingness, overlap and code-pin checks. These counts describe computational checks, not independent statistical replications or proof of causal identification. Raw DBC payloads and some large primary-source files are not included in the compact replay; direct hash verification does not independently decode or re-extract them. Access failures, conflicting years/geographies and unavailable cells retain explicit dispositions. The source-only companions above add zero effect-register records and zero independent likelihood weight. Their value is fuller measurement and provenance accounting, not artificial tightening of the main estimate.

S73.6 Source-handling and measurement checks

The supplied source-audit companions distinguish strings, identifiers, missingness, suppression, date labels and denominator universes before analysis. South Carolina age/date checks and Connecticut identifiers, units and rounded-share arithmetic complement historical-risk and power-table checks. British Columbia and Norway retain pinned comparison receipts; compact verification does not repeat their original raw comparisons. The immutable contracts prevent a moving report file from being mistaken for a historical execution input.

A bounded header review has 406 candidates: 332 supported, 67 unresolved and seven ambiguous. The 237 overrides are metadata dispositions, not silently inferred vaccine exposures. Discovery uses bounded-prefix hashes, not full-file certification. South Carolina's 95 checked single-age denominator cells distinguish resident from civilian population and school enrollment; an earlier source trace is pinned rather than repeated by compact replay.

Australian NDIS checks cover five files representing three unique snapshots and retain 27 diagnosis-overlap envelopes. A displayed `<N` allows nonnegative integer values zero through N-1; it is not a statistical confidence interval. Nonexclusive reported diagnoses cannot be summed as mutually exclusive disability counts, consistent with the NDIA diagnosis rules. New Jersey retains 81 omitted blank-county charter rows and 831 genuine reported zero cells as distinct dispositions. Twelve source-count partitions and 228 unchanged aggregation cells are checked; missing county or identifier fields are not imputed.

New York's selected-source audit checks 117 HTML payloads, excludes 14 stale payloads and supplies 740 additive legacy-date corrections (682 district rows and 58 county totals). The current model panel is not silently replaced by an alternative extraction vintage. Displayed zeros remain censoring-ambiguous, signed residuals down to -10 remain signed, and the 2012 reporting break remains unresolved. Its model sensitivity is reported in S71.8 rather than treated as a source correction that establishes causal comparability.

Source-only checks add no coefficients or independent likelihood weight. Compact mode verifies supplied hashes, missingness and arithmetic; the optional full mode additionally hashes named omitted originals and bounded prefixes, without re-parsing original workbooks or HTML, decoding large source payloads, refitting models or executing native Excel. Neither tier certifies worldwide source completeness. Exact scopes and pins are supplied in `qa/data_handling_20261002/acquired_audit_integration_3665/v2/README.md` and `source_handling_closure_3684/REVIEW.md`; age-model replay has its separate contract described in S71.7.

Appendix Table SA1. Machine-linked robustness catalog

Notes: Pass/fail denotes the listed diagnostic criterion, not causal validity or absence of harm. Cross-condition outcomes are not assumed to be negative controls; non-reject means that the corresponding coefficient test did not reject. Uptake checks do not validate the residualized PCV7 exposure mapping (S52).

Audit ID	Design	Family	Test	Result	p	Status
0001	United States	positive control	PCV7 introduction	34.828	<0.001	pass
0002	United States	positive control	New rotavirus vaccine introduction	18.442	<0.001	pass
0003	U.S. SchoolVaxView	positive control	Exemption-to-MMR measurement check	-1.165	0.129	lim./fail
0004	Germany	positive control	PCV complete series	68.410	-	pass
0005	Germany	positive control	Meningococcal C	72.890	-	pass
0006	Germany	positive control	Varicella 1+ dose	43.810	-	pass
0007	Germany	positive control	Varicella 2 doses	63.520	-	pass
0008	Italy	positive control	Varicella nationalization	16.270	-	pass
0009	Italy	positive control	Rotavirus nationalization	36.650	-	pass
0010	Sweden	positive control	PCV programme uptake	55.167	-	pass
0011	Sweden	positive control	HEPB programme uptake	97.276	-	pass
0012	Sweden	positive control	ROTAVIRUS programme uptake	84.019	-	pass
0013	U.S. IDEA PCV7	assignment placebo	State-intensity permutation	0.019	0.333	pass
0014	U.S. IDEA PCV7	pretrend placebo	Joint event-study leads	-	0.636	pass
0015	SchoolVaxView MMR	assignment placebo	Within-cohort state permutation	-0.019	0.251	pass
0016	SchoolVaxView MMR	future-exposure placebo	Future coverage	0.053	0.287	pass
0017	ADDM PCV7	assignment placebo	Site-intensity permutation	0.052	0.409	pass
0018	England rotavirus	timing placebo	False cutoff 2004	-0.018	0.318	pass
0019	England rotavirus	timing placebo	False cutoff 2006	-0.018	0.164	pass
0020	England rotavirus	timing placebo	False cutoff 2008	0.022	0.104	pass
0021	England rotavirus	timing placebo	False cutoff 2010	0.002	0.922	pass
0022	Wales rotavirus	assignment placebo	Authority-uptake permutation	-0.133	0.598	pass
0023	Wales rotavirus	assignment placebo	Interaction permutation	-0.015	0.925	pass
0024	Northern Ireland rotavirus	timing placebo	Placebo-cutoff rank	-	0.875	pass
0025	Yokohama MMR withdrawal	timing placebo	Withdrawal cutoff 1990	0.135	0.739	pass
0026	Yokohama MMR withdrawal	timing placebo	Withdrawal cutoff 1991	-0.279	0.470	pass
0027	Yokohama MMR withdrawal	timing placebo	Withdrawal cutoff 1993	0.554	0.105	pass
0028	Yokohama MMR withdrawal	timing placebo	Withdrawal cutoff 1995	-0.405	0.133	pass
0029	Scotland rotavirus	pretrend placebo	Joint pretrend	-	0.137	pass
0030	Wales rotavirus	pretrend placebo	Joint pretrend	-	0.010	fail
0031	Spain varicella	pretrend placebo	Joint pretrend	-	0.020	fail
0032	U.S. IDEA PCV7	cross-condition outcome	Autism	1.013	0.419	target
0033	U.S. IDEA PCV7	cross-condition outcome	Hearing impairments	0.968	0.381	non-reject
0034	U.S. IDEA PCV7	cross-condition outcome	Visual impairments	0.837	0.090	non-reject
0035	U.S. IDEA PCV7	cross-condition outcome	Orthopedic impairments	1.023	0.602	non-reject
0036	U.S. IDEA PCV7	cross-condition outcome	Traumatic brain injury	1.032	0.680	non-reject
0037	U.S. IDEA PCV7	cross-condition outcome	Speech or language impairments	0.995	0.785	non-reject
0038	U.S. IDEA PCV7	cross-condition outcome	Intellectual disability	1.028	0.360	non-reject
0039	U.S. IDEA PCV7	cross-condition outcome	Other health impairments	1.025	0.294	non-reject
0040	Scotland rotavirus	cross-condition outcome	Autism	1.020	0.946	target
0041	Scotland rotavirus	cross-condition outcome	Hearing	0.566	0.200	non-reject
0042	Scotland rotavirus	cross-condition outcome	Visual	0.641	0.342	non-reject
0043	Scotland rotavirus	cross-condition outcome	Physical Motor	0.835	0.492	non-reject
0044	Scotland rotavirus	cross-condition outcome	Speech Language	0.723	0.119	non-reject
0045	Scotland rotavirus	cross-condition outcome	Learning Disability	0.581	0.083	non-reject
0046	Wales rotavirus	cross-condition outcome	Autism	0.875	0.639	target
0047	Wales rotavirus	cross-condition outcome	Physical Medical	0.662	0.638	non-reject
0048	Wales rotavirus	cross-condition outcome	Hearing	0.811	0.820	non-reject
0049	Wales rotavirus	cross-condition outcome	Visual	0.944	0.940	non-reject
0050	Wales rotavirus	cross-condition outcome	Speech Language	0.881	0.888	non-reject

Audit ID	Design	Family	Test	Result	p	Status
0051	Wales rotavirus	cross-condition outcome	Moderate Learning	0.241	0.310	non-reject
0052	Sweden PCV	cross-condition outcome	Autism category (F84)	0.886	0.292	target
0053	Sweden PCV	cross-condition outcome	Speech/language (F80)	1.012	0.939	non-reject
0054	Sweden PCV	cross-condition outcome	Motor development (F82)	1.062	0.778	non-reject
0055	Sweden PCV	cross-condition outcome	Hyperkinetic disorders (F90)	0.742	0.019	non-reject
0056	U.S. IDEA PCV7	alternative estimator/specification	PPML, state/cohort FE	1.017	0.368	estimate
0057	U.S. IDEA PCV7	alternative estimator/specification	Population-weighted log-rate WLS	1.019	0.458	estimate
0058	U.S. IDEA PCV7	alternative estimator/specification	Unweighted log-rate OLS	0.988	0.617	estimate
0059	U.S. IDEA PCV7	alternative estimator/specification	GEE Poisson, independent	1.017	0.321	estimate
0060	U.S. IDEA PCV7	alternative estimator/specification	GEE Poisson, exchangeable	1.017	0.323	estimate
0061	U.S. IDEA PCV7	alternative estimator/specification	WLS plus intensity-specific linear trend	1.023	0.234	estimate
0062	U.S. IDEA PCV7	alternative estimator/specification	WLS plus linear and quadratic intensity trends	1.022	0.268	estimate
0063	U.S. IDEA PCV7	alternative estimator/specification	WLS, endpoint cohorts removed	1.018	0.483	estimate
0064	SchoolVaxView MMR	alternative estimator/specification	PPML, state/cohort FE	0.989	0.718	estimate
0065	SchoolVaxView MMR	alternative estimator/specification	Population-weighted log-rate WLS	0.981	0.498	estimate
0066	SchoolVaxView MMR	alternative estimator/specification	Unweighted log-rate OLS	0.983	0.532	estimate
0067	SchoolVaxView MMR	alternative estimator/specification	GEE Poisson, independent	0.989	0.646	estimate
0068	SchoolVaxView MMR	alternative estimator/specification	GEE Poisson, exchangeable	0.989	0.646	estimate
0069	SchoolVaxView MMR	alternative estimator/specification	WLS with state-specific cohort trends	0.959	0.531	estimate
0070	England rotavirus	alternative estimator/specification	PPML with flexible cohort spline	0.949	0.022	estimate
0071	England rotavirus	alternative estimator/specification	Population-weighted log-rate WLS	0.950	0.029	estimate
0072	England rotavirus	alternative estimator/specification	Unweighted log-rate OLS	0.948	0.028	estimate
0073	England rotavirus	alternative estimator/specification	GEE Poisson, exchangeable	0.937	0.011	estimate
0074	England rotavirus	alternative estimator/specification	PPML plus SEN-reform indicator	0.948	0.021	estimate
0075	Scotland rotavirus	alternative estimator/specification	PPML with uptake-specific spline	1.020	0.946	estimate
0076	Scotland rotavirus	alternative estimator/specification	Population-weighted log-rate WLS	0.876	0.603	estimate
0077	Scotland rotavirus	alternative estimator/specification	Unweighted log-rate OLS	0.844	0.464	estimate
0078	Scotland rotavirus	alternative estimator/specification	GEE Poisson, exchangeable	1.020	0.940	estimate
0079	Wales rotavirus	alternative estimator/specification	PPML with uptake-specific spline	0.875	0.639	estimate
0080	Wales rotavirus	alternative estimator/specification	Population-weighted log-rate WLS	0.901	0.705	estimate
0081	Wales rotavirus	alternative estimator/specification	Unweighted log-rate OLS	0.771	0.386	estimate
0082	Wales rotavirus	alternative estimator/specification	GEE Poisson, exchangeable	0.876	0.602	estimate
0083	Wales rotavirus	alternative estimator/specification	PPML plus ALN-system intensity break	1.051	0.652	estimate
0084	ADDM PCV7	alternative estimator/specification	Site and cohort fixed effects	1.072	0.190	estimate
0085	ADDM PCV7	alternative estimator/specification	Intensity-specific linear trend	0.947	0.368	estimate

Audit ID	Design	Family	Test	Result	p	Status
0086	ADDM PCV7	alternative estimator/specification	Intensity-specific flexible spline	1.053	0.163	estimate
0087	U.S. IDEA PCV7	design audit	positive_control	-	-	pass
0088	U.S. IDEA PCV7	design audit	negative_outcomes	-	-	pass
0089	U.S. IDEA PCV7	design audit	placebo_or_pretrend	-	-	pass
0090	U.S. IDEA PCV7	design audit	trend_variants	-	-	pass
0091	U.S. IDEA PCV7	design audit	alternative_estimators	-	-	pass
0092	U.S. IDEA PCV7	design audit	influence_or_randomization	-	-	pass
0093	U.S. IDEA PCV7	design audit	moderate_effect_power	-	-	Adequate
0094	PCV7 diagnostic-substitution DDD	design audit	positive_control	-	-	pass
0095	PCV7 diagnostic-substitution DDD	design audit	negative_outcomes	-	-	pass
0096	PCV7 diagnostic-substitution DDD	design audit	placebo_or_pretrend	-	-	pass
0097	PCV7 diagnostic-substitution DDD	design audit	trend_variants	-	-	pass
0098	PCV7 diagnostic-substitution DDD	design audit	alternative_estimators	-	-	pass
0099	PCV7 diagnostic-substitution DDD	design audit	influence_or_randomization	-	-	pass
0100	PCV7 diagnostic-substitution DDD	design audit	moderate_effect_power	-	-	Adequate
0101	SchoolVaxView MMR	design audit	positive_control	-	-	limited
0102	SchoolVaxView MMR	design audit	negative_outcomes	-	-	pass
0103	SchoolVaxView MMR	design audit	placebo_or_pretrend	-	-	pass
0104	SchoolVaxView MMR	design audit	trend_variants	-	-	limited
0105	SchoolVaxView MMR	design audit	alternative_estimators	-	-	pass
0106	SchoolVaxView MMR	design audit	influence_or_randomization	-	-	pass
0107	SchoolVaxView MMR	design audit	moderate_effect_power	-	-	Adequate
0108	ADDM PCV7	design audit	positive_control	-	-	pass
0109	ADDM PCV7	design audit	negative_outcomes	-	-	limited
0110	ADDM PCV7	design audit	placebo_or_pretrend	-	-	pass
0111	ADDM PCV7	design audit	trend_variants	-	-	limited
0112	ADDM PCV7	design audit	alternative_estimators	-	-	pass
0113	ADDM PCV7	design audit	influence_or_randomization	-	-	pass
0114	ADDM PCV7	design audit	moderate_effect_power	-	-	Limited
0115	England rotavirus	design audit	positive_control	-	-	pass
0116	England rotavirus	design audit	negative_outcomes	-	-	pass
0117	England rotavirus	design audit	placebo_or_pretrend	-	-	pass
0118	England rotavirus	design audit	trend_variants	-	-	pass
0119	England rotavirus	design audit	alternative_estimators	-	-	pass
0120	England rotavirus	design audit	influence_or_randomization	-	-	pass
0121	England rotavirus	design audit	moderate_effect_power	-	-	Adequate
0122	Scotland rotavirus	design audit	positive_control	-	-	pass
0123	Scotland rotavirus	design audit	negative_outcomes	-	-	pass
0124	Scotland rotavirus	design audit	placebo_or_pretrend	-	-	pass
0125	Scotland rotavirus	design audit	trend_variants	-	-	pass
0126	Scotland rotavirus	design audit	alternative_estimators	-	-	limited
0127	Scotland rotavirus	design audit	influence_or_randomization	-	-	pass
0128	Scotland rotavirus	design audit	moderate_effect_power	-	-	Low
0129	Wales rotavirus	design audit	positive_control	-	-	pass
0130	Wales rotavirus	design audit	negative_outcomes	-	-	pass
0131	Wales rotavirus	design audit	placebo_or_pretrend	-	-	fail

Audit ID	Design	Family	Test	Result	p	Status
0132	Wales rotavirus	design audit	trend_variants	-	-	limited
0133	Wales rotavirus	design audit	alternative_estimators	-	-	pass
0134	Wales rotavirus	design audit	influence_or_randomization	-	-	pass
0135	Wales rotavirus	design audit	moderate_effect_power	-	-	Low
0136	Northern Ireland rotavirus	design audit	positive_control	-	-	pass
0137	Northern Ireland rotavirus	design audit	negative_outcomes	-	-	limited
0138	Northern Ireland rotavirus	design audit	placebo_or_pretrend	-	-	fail
0139	Northern Ireland rotavirus	design audit	trend_variants	-	-	limited
0140	Northern Ireland rotavirus	design audit	alternative_estimators	-	-	limited
0141	Northern Ireland rotavirus	design audit	influence_or_randomization	-	-	pass
0142	Northern Ireland rotavirus	design audit	moderate_effect_power	-	-	Low
0143	Sweden PCV	design audit	positive_control	-	-	pass
0144	Sweden PCV	design audit	negative_outcomes	-	-	pass
0145	Sweden PCV	design audit	placebo_or_pretrend	-	-	fail
0146	Sweden PCV	design audit	trend_variants	-	-	limited
0147	Sweden PCV	design audit	alternative_estimators	-	-	limited
0148	Sweden PCV	design audit	influence_or_randomization	-	-	pass
0149	Sweden PCV	design audit	moderate_effect_power	-	-	Low
0150	Spain varicella	design audit	positive_control	-	-	pass
0151	Spain varicella	design audit	negative_outcomes	-	-	limited
0152	Spain varicella	design audit	placebo_or_pretrend	-	-	fail
0153	Spain varicella	design audit	trend_variants	-	-	pass
0154	Spain varicella	design audit	alternative_estimators	-	-	limited
0155	Spain varicella	design audit	influence_or_randomization	-	-	pass
0156	Spain varicella	design audit	moderate_effect_power	-	-	Low
0157	Yokohama MMR withdrawal	design audit	positive_control	-	-	pass
0158	Yokohama MMR withdrawal	design audit	negative_outcomes	-	-	limited
0159	Yokohama MMR withdrawal	design audit	placebo_or_pretrend	-	-	pass
0160	Yokohama MMR withdrawal	design audit	trend_variants	-	-	limited
0161	Yokohama MMR withdrawal	design audit	alternative_estimators	-	-	limited
0162	Yokohama MMR withdrawal	design audit	influence_or_randomization	-	-	pass
0163	Yokohama MMR withdrawal	design audit	moderate_effect_power	-	-	Low
0164	U.S. RotaShield pulse	design audit	positive_control	-	-	pass
0165	U.S. RotaShield pulse	design audit	negative_outcomes	-	-	limited
0166	U.S. RotaShield pulse	design audit	placebo_or_pretrend	-	-	limited
0167	U.S. RotaShield pulse	design audit	trend_variants	-	-	pass
0168	U.S. RotaShield pulse	design audit	alternative_estimators	-	-	limited
0169	U.S. RotaShield pulse	design audit	influence_or_randomization	-	-	pass
0170	U.S. RotaShield pulse	design audit	moderate_effect_power	-	-	Adequate

Notes: Source: project_results/scotland_support/corrected_robustness_catalog.csv, derived from 134_robustness_complete_catalog.csv with source-verified hearing estimates (S57). Negative-control-null statuses are displayed as non-rejection. PDF IDs omit the repeated ROBUST- prefix. Audit identifiers retain unfavorable as well as favorable tests.

S74. Aggregate event-study methods and reference analyses

This section preserves the documented methods and fourth-edition reference analyses of the event-study pipeline. Subsequent source expansions and the public v5e/A7 results are documented in Sections S77-S79. The main paper reports the original A7 sample: 39,753 comparisons, 404 changes and 40 countries. Historical rows and artwork retain their own stated editions; they are alternative analyses of overlapping data, not independent confirmations. Source-ledger records establish provenance and do not establish causal validity.

S74.1 Versions and their rules

Versions 1-4 were each written as a rule and hashed before any of their fits; later changes are listed as post hoc amendments beside the rule. The v5 rule (RULE 9160) was pinned by SHA-256 in an execution freeze that Codex wrote (freeze 4977) before the first model call (10 October 2026, 22:03 CDT); the companion hash files the rule promised were not written. The rule was written after earlier versions' results had been seen, so v5 is a disclosed update, not a preregistered analysis. Version 5a was fitted by Codex with the same code (continuation 4980); the cited pool, 5083, adds Sweden's corrected source counts; version 5c entered through a specified amendment (A2) written before its fits. Amendment A2 also made series observed in fewer than four distinct calendar years inside the estimation window sensitivity-only; the rule was prompted by planted-effect diagnostics on preview data whose estimates had been seen, and it moved series from Israel, Italy, New Zealand and Tunisia to a sensitivity that gives 0.975 (placebo-calibrated 0.858-1.107). Version 1 (RULE_9150) extrapolated each jurisdiction's pre-change trend; version 2 (RULE_9157) fitted the trend and the step jointly and modelled age-band series; version 3 (RULE_9158, with RULE_9158b and RULE_9158c for Brazil's school census) added the October country expansion; version 4 (RULE_9159) pools countries by random effects, adds measured U.S. sub-state coverage and Brazil's outpatient records, and was hashed before any v4 fit. Two runs of the v4 pooling were identical.

Table S74-A. Per-child rate ratios across versions. Intervals in versions 1-3 are data-estimated (placebo) 95 percent intervals under fixed-information weights; version 4 reports the placebo-calibrated interval under random-effects weights. Effective K is the effective number of countries.

Version	Weighting	Countries	Per-child relative risk	95% interval	Effective K
v1 (RULE_9150)	fixed information	7	0.954	0.686-1.328	2.10
v2 (RULE_9157)	fixed information	14	0.949	0.633-1.425	3.01
v3 final, no school census (RULE_9158)	fixed information	28	0.929	0.643-1.406	3.91
v5c data, v3 weighting (sensitivity)	fixed information	40	0.967	0.739-1.265	3.60
v4 (RULE_9159)	random effects by country	27	0.965	0.810-1.150	18.76
v5a (RULE_9160, Codex pool 5083)	random effects by country	32	0.963	0.820-1.130	22.38
v5c with A3 (RULE_9160 with amendments A2 and A3)	random effects by country	37	0.973	0.834-1.136	24.90
v5e headline (RULE_9160 with amendments A2-A4 and A7)	random effects by country	40	0.973	0.836-1.131	26.04

The post hoc amendments to RULE_9159 change no specified estimate, gate or interval. P1 added planted-effect fits for the spline specification, which had topped the specified precision ranking. P2 edited figure labels. P3 recorded readings: the kindergarten-measured school-entry panel is not validated, the near-empty blank-diagnosis subset of Brazil's outpatient records is uninformative, and Italy is excluded on the grid boundary. P4, a coordinator decision, identified the most precise specification passing both the planted-effect and placebo gates. P5, by the author's decision of 10 October, redrew every figure and the ramp test from the headline specification, which every number in the papers and the post uses; the most precise gated version is shown here (Figure S119), and the charts count comparisons after the boundary exclusions.

S74.2 Data layer

The harmonized data layer registers 272 sources. Its primary autism data hold 3,145,790 cells for 29,170 jurisdictions in 33 countries. Measured U.S. sub-state coverage adds 13,220,271 rows from 108 source parts, at the district, intermediate-unit and county levels, each tagged as infant, child-care or kindergarten coverage and as measured or crosswalked. The event register holds 3,960 events, of which 3,905 are primary after adjudication: statute-cited exemption-law events supersede the duplicate register rows for the same state and year, and four events supported only by general knowledge (Czech BCG withdrawal 2010, German MCV2 mandate 2020, French BCG withdrawal 2007 and French MenB introduction 2022) stay out of the primary until verified. Of the exemption-law events, 17 were infant-relevant in version 4 and enter the primary at the first birth cohort entering child care after the law; the other ten change only school-entry requirements and form the negative-control event set (S74.5). Two consecutive builds of the data layer were deterministic.

One limitation of the coverage layer remains. Exemption rates measured in different settings of the same state source (child care, kindergarten and all grades) share an extended key, so 112,408 keys collide across sibling files. Version 4 never uses those collided exemption rows; where exemption rates enter (only the kindergarten-measured negative-control panel), they are selected by setting-specific source. A permanent fix needs a setting field in the key.

S74.3 Design and estimator

The outcome-blind design has 95,462 jurisdiction-by-change comparisons in 11,169 series, 41 countries; 84,524 comparisons have a known uptake change. For each comparison, a quasi-Poisson model is fitted jointly to birth cohorts six before to six after the first affected cohort, with age or band effects, one cohort trend through the window (linear, quadratic or restricted cubic spline) and a step. The raw step is an indicator for affected cohorts. The per-child step is theta times the cohort's measured uptake change p, the linearized form of log[1 + p(r - 1)], with r = exp(theta). Dispersion is estimated from cohort-grouped Pearson residuals with a floor of one. Within each country the step is the maximum of the summed profile likelihoods, with variance equal to the larger of the jurisdiction-clustered and model-based variances; each country's variance is then widened to cover the sampling intervals of the three trend shapes.

Across countries, weights are $1/(v + \tau^2)$ with $\tau = 0.33$ on the log scale, the between-country standard deviation estimated from the version 3 placebo dates and held fixed in this analysis. The random-effects interval uses variance 1/(sum of weights). The placebo-calibrated interval re-estimates the bias components from version 4's own placebo fits, run at six fake dates seven to twelve cohorts before each real change: the shared component is 0.00 (profile interval upper end 0.08) and the country component 0.260 (0.23-0.295), estimated from 187 country-by-offset placebo steps in 34 countries. The outer band adds an assumed shared allowance of 0.15. Weights stay at $\tau = 0.33$ in every interval.

Supplement Figures S114 and S115 illustrate one comparison and the pooling from comparisons to one number.

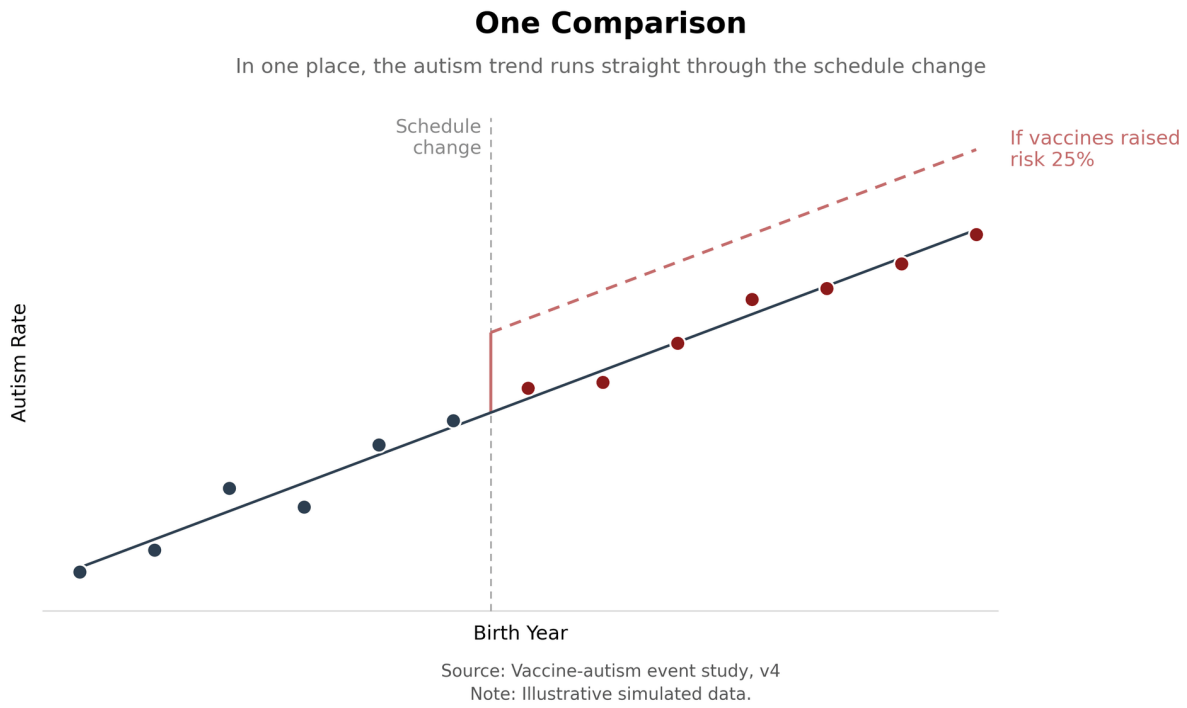


Figure S114. One comparison. In one jurisdiction, autism rates by birth year follow one continuous trend through a schedule change; the dashed line shows how far a per-child relative risk of 1.25 would have lifted the cohorts born after it. Illustrative simulated data.

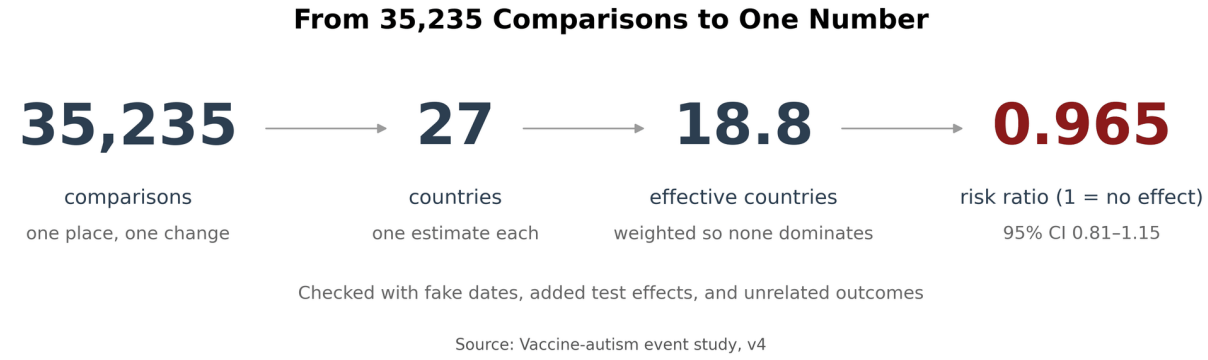


Figure S115. Historical fourth-edition pooling diagram: 35,235 comparisons are combined within 27 countries, giving an estimate of 0.965. These are the numbers printed in the retained reference diagram, not the current main-paper sample. The diagram illustrates within-country and then between-country pooling; a bias shared across countries does not average down.

S74.4 Results

The headline, a linear trend on all primary comparisons, pools 39,753 comparisons in 40 countries (404 distinct schedule changes): per-child relative risk 0.973, random-effects 95 percent CI 0.836-1.131, placebo-calibrated 0.855-1.107, outer band 0.699-1.354, effective K 26.04. The raw step is +1.7 percent (random-effects -7.0 to +11.3 percent; placebo-calibrated -5.9 to +9.9 percent) across 46,127 comparisons in 41 countries. Brazil, the United Kingdom and the United States each carry about 5 percent of the per-child weight; under the version 3 weighting the United Kingdom and the United States together carried 70.1 percent (Figure S116).

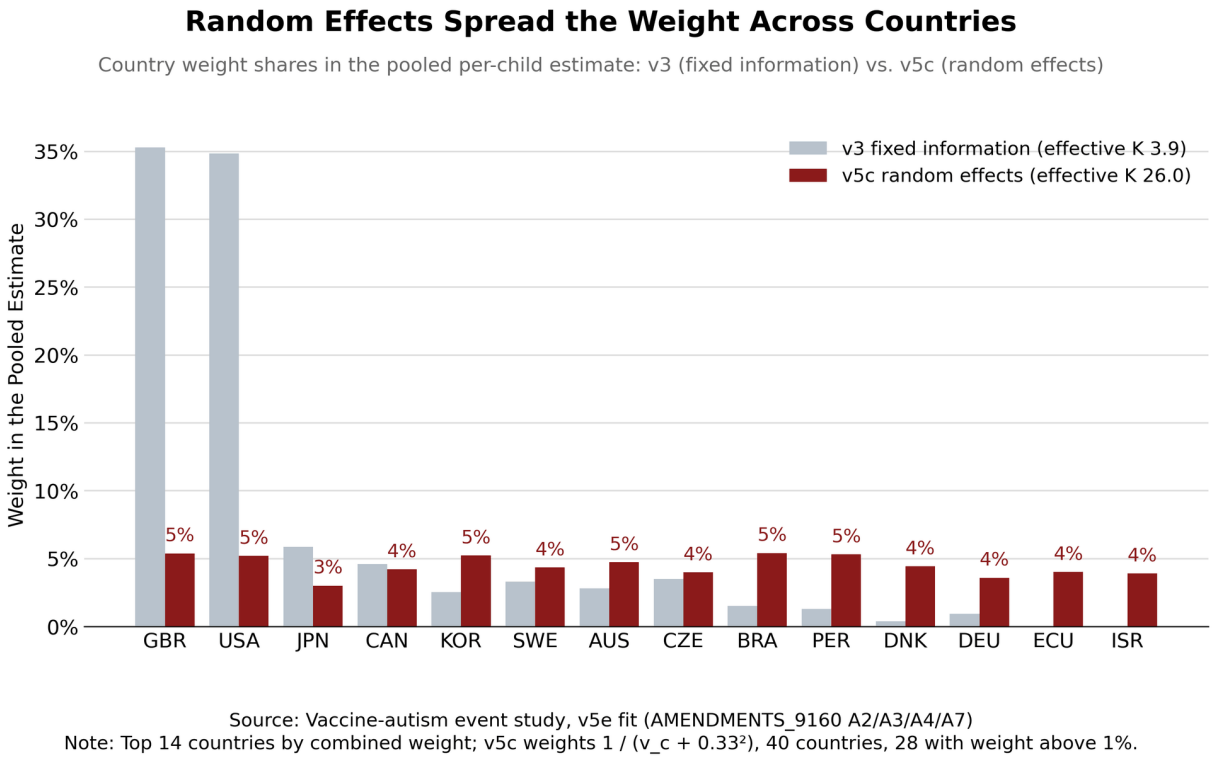


Figure S116. Country weight shares in the per-child estimate of the aggregate event study: version 3 (one profile likelihood with fixed-information weights; effective number of countries 3.91) against the headline, version 5e (random effects by country, weights 1/(v + 0.33^2); effective number 26.04). The countries with the largest combined weight are shown.

In version 4, of 52 specifications (trend shape by window by subset), 20 pass both the planted-effect gate and the placebo gate, and all 52 pass the placebo gate; the headline passes both. The most precise passing specification, a quadratic trend without mandate events, gives 0.948 (placebo-calibrated 0.800-1.124) from 29,531 comparisons and 171 distinct changes in 26 countries, effective K 19.20 (Figure S119). The spline trend on all comparisons gives 1.014 (0.857-1.200) but recovers a planted 1.25 as 1.20 and fails the gate. On all comparisons only the linear trend passes: the quadratic trend and the window ending four cohorts after the change miss the gate narrowly (recovery errors -10.1 and -10.4 percent).

Table S74-B. Sensitivities (per-child relative risk, linear trend unless stated). Calibrated intervals use the v4 placebo components; random-effects intervals use tau 0.33 or the stated value.

Specification	RR	Calibrated 95% CI	Random-effects 95% CI	Effective K
Headline: all comparisons	0.973	0.855-1.107	0.836-1.131	26.04
Quadratic trend (misses planted gate narrowly)	0.960	0.846-1.088	0.828-1.112	26.76
Spline trend (fails planted gate)	0.980	0.866-1.109	0.848-1.133	27.26
Window -6 to +4 (misses planted gate narrowly)	0.963	0.848-1.093	0.829-1.118	26.15
No mandate events	0.970	0.852-1.105	0.834-1.129	26.11
Without Brazil outpatient records (AM-1 veto)	0.974	0.856-1.109	0.837-1.133	26.09
Between-country SD 0.285	0.972	0.855-1.106	0.848-1.115	24.78
Between-country SD 0.385	0.973	0.853-1.109	0.822-1.152	27.26
Between-country SD from v4 placebo dates	0.972	0.855-1.105	0.855-1.105	23.93
Version 3 weighting (one profile likelihood)	0.967	0.739-1.265	0.954-0.979	3.60
U.S. kindergarten proxies added	0.973	0.855-1.107	0.836-1.131	26.04
Brazil outpatient, treated cases	0.974	0.855-1.108	0.837-1.132	26.03
Brazil outpatient, treated cases, low-blank state-years	0.975	0.856-1.109	0.838-1.134	26.04
German series by place of treatment	0.974	0.856-1.108	0.837-1.133	26.04

Specification	RR	Calibrated 95% CI	Random-effects 95% CI	Effective K
Low voluntary-uptake scenario (Japan, Korea, Taiwan)	0.970	0.852-1.104	0.834-1.128	26.09
High voluntary-uptake scenario (Japan, Korea, Taiwan)	0.976	0.857-1.112	0.839-1.136	25.91

In the version 3 weighting row the second interval is that method's sampling-only interval, which has no between-country term; its calibrated interval is the comparable one.

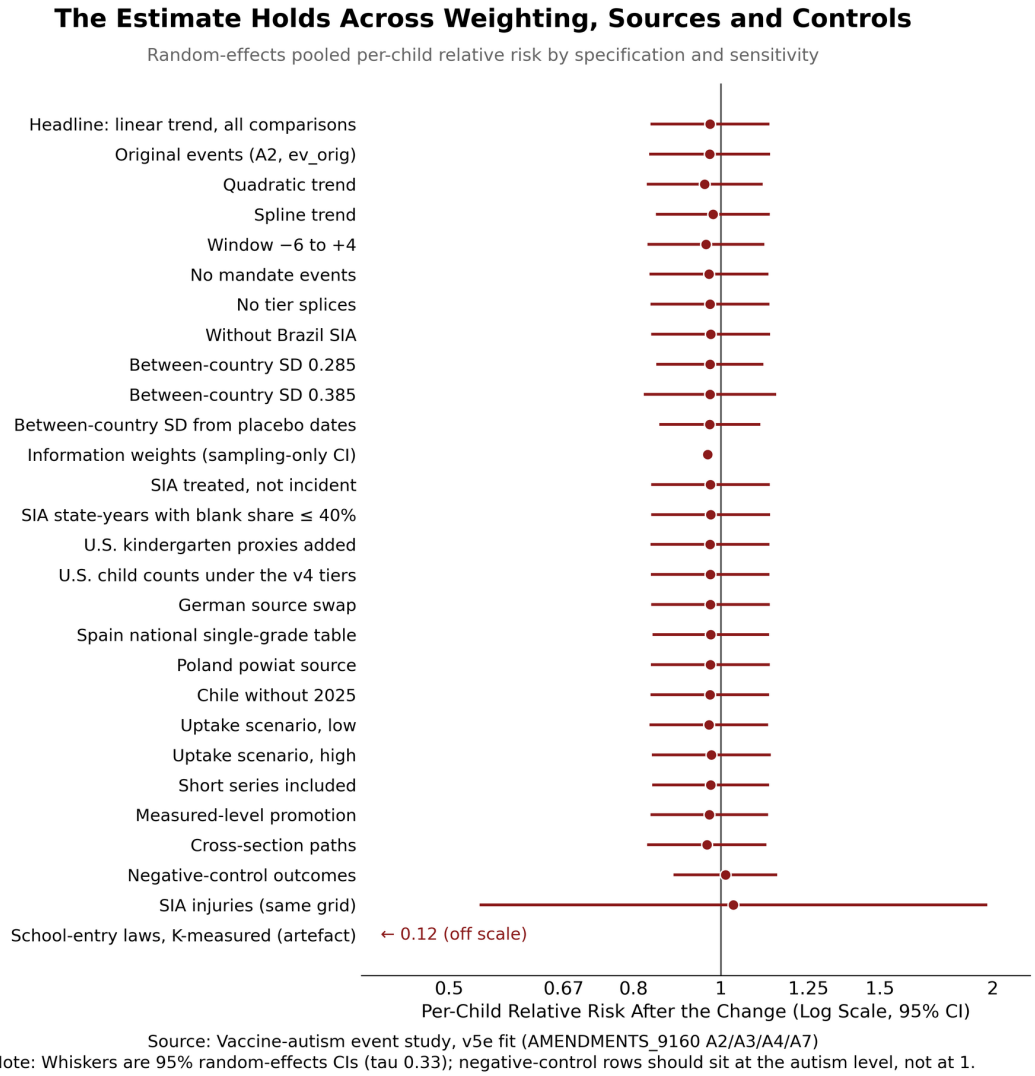
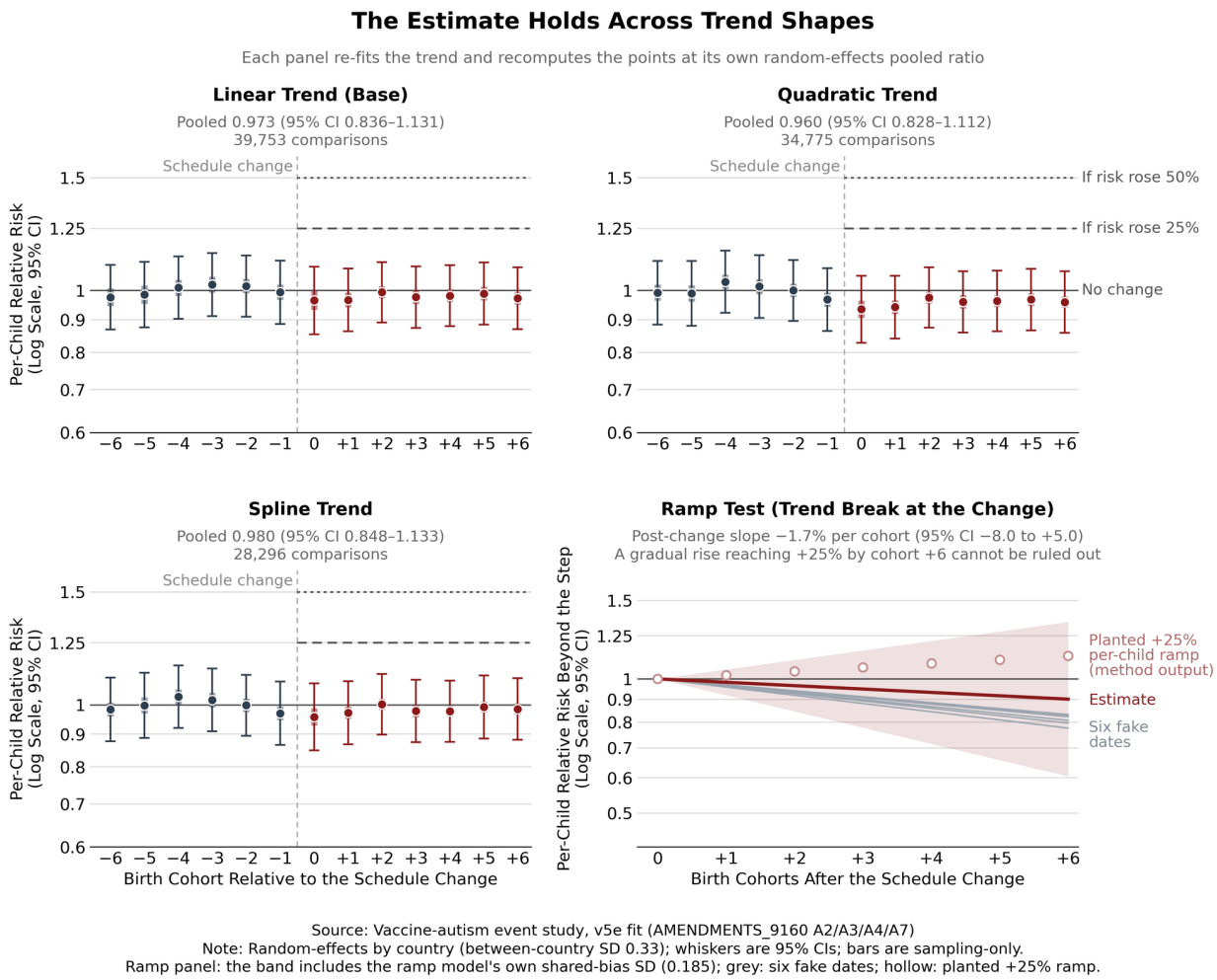


Figure S117. Sensitivities of the aggregate event study (version 5e). Random-effects pooled per-child relative risks with random-effects 95 percent intervals, by event set, trend shape, window, between-country standard deviation, weighting, source, uptake scenario and negative-control panel; the information-weights row shows a sampling-only interval. The kindergarten-measured school-entry panel is an artefact of a near-zero uptake change and carries no claim. Negative-control outcome rows are to be compared with the autism estimate, not with one.

The trend shapes and the ramp test are drawn in Figure S118, and the most precise gated version in Figure S119.



Source: Vaccine-autism event study, v5e fit (AMENDMENTS_9160 A2/A3/A4/A7)
Note: Random-effects by country (between-country SD 0.33); whiskers are 95% CIs; bars are sampling-only.
Ramp panel: the band includes the ramp model's own shared-bias SD (0.185); grey: six fake dates; hollow: planted +25% ramp.

Figure S118. Trend shapes and the ramp test for the headline specification (version 5e). Each panel refits the trend on all comparisons and pools at its own random-effects step, with random-effects 95 percent intervals: linear 0.973 (0.836-1.131), quadratic 0.960 (0.828-1.112) and spline 0.980 (0.848-1.133). The ramp panel, fitted on 35,365 comparisons, lets the per-child effect grow after the change: slope -1.73 percent per cohort (placebo-calibrated -8.05 to +5.02); a planted ramp reaching 25 percent by the sixth cohort is recovered with an error of +0.5 percent, and the real-data interval does not exclude a ramp of that size.

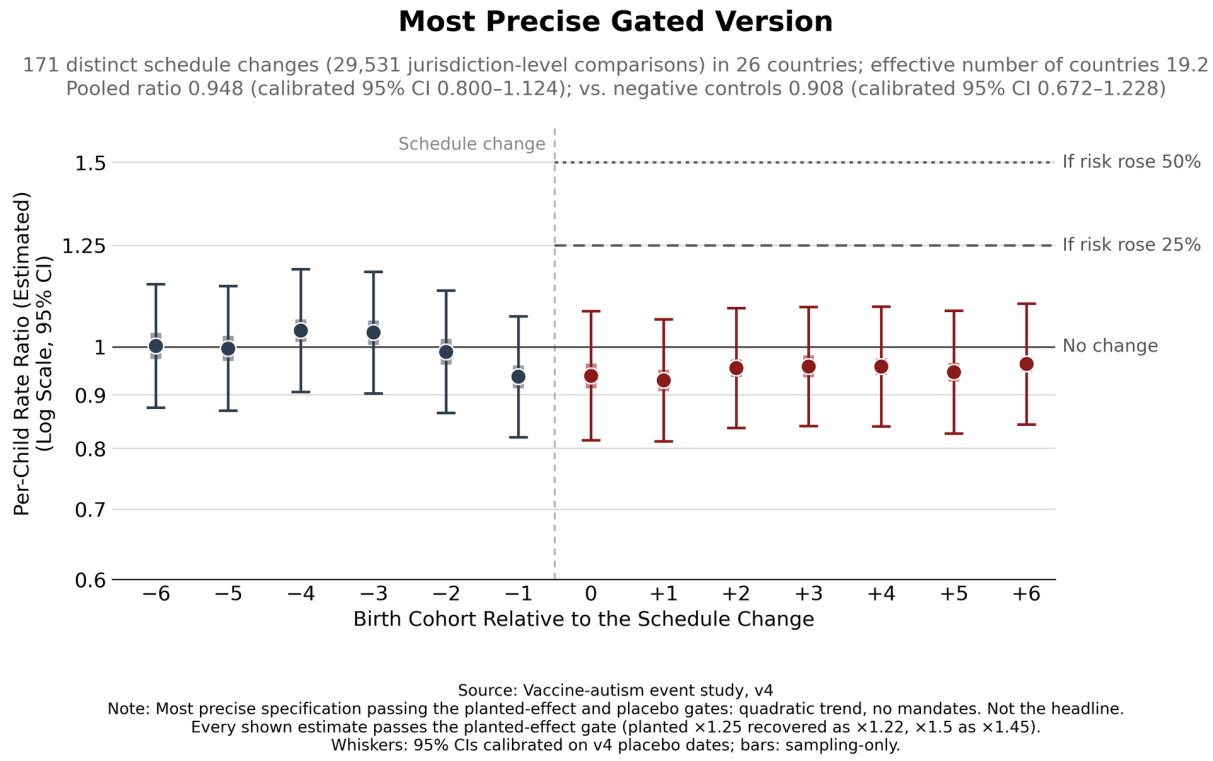


Figure S119. Version 4 run: The most precise gated version: a quadratic trend without mandate events, the most precise of the 20 specifications that pass both the planted-effect and placebo gates. 29,531 comparisons and 171 distinct changes in 26 countries (effective number 19.2); pooled step 0.948 (placebo-calibrated 0.800-1.124). A sensitivity of the headline, not a replacement for it.

S74.5 Gates

Planted effects. Injecting a per-child relative risk of 1.25 or 1.5 into the real post-change cells, along each comparison's uptake path (measured or modeled), and rerunning the whole pipeline with the real country weights returns 1.23 and 1.46: log-scale errors of -8.3 and -6.7 percent, inside the 10 percent gate. On the raw scale the errors are -10.6 and -14.5 percent; the raw version has no planted-effect gate. In version 4 every spline specification failed, recovering 1.25 as about 1.20, and on all comparisons the quadratic trend missed the gate narrowly (-10.1 percent). The estimator recovers about 92 percent of a planted log-step in version 5c (attenuation toward one).

Placebo dates. All six random-effects pooled placebo steps (offsets seven to twelve) lie inside their random-effects intervals, with a largest absolute z of 1.20. Country-by-offset placebo estimates fall inside their sampling-only intervals 0.80 of the time (187 estimates), which is why the calibrated interval adds the estimated country component.

Negative-control outcomes. Outcomes vaccines should not affect (sensory and physical disability, chronic illness, injuries and accidents, appendicitis, Down syndrome and ringworm) give a pooled per-child ratio of 1.011 (calibrated 0.908-1.127). Autism relative to the controls is 0.962 (calibrated 0.813-1.138). Brazilian injury records on the outpatient grid give 1.033 (0.620-1.719).

Negative-control events. School-entry-only exemption laws change compliance at school entry, not infant receipt, so autism should not step at them. On the kindergarten-measured panel the raw step is +2.0 percent (calibrated -22.9 to +34.9) on 37 comparisons, and on state series +5.4 percent (-36.5 to +75.1) on 10 comparisons; both include zero. Only Pennsylvania's 2017 law has measured kindergarten paths (28 intermediate units, mean first stage +0.034). On that panel planted effects are recovered with errors of -34.6 and -28.4 percent, so the panel is not validated and its per-child estimate, 0.123 (0.051-0.294), carries no claim.

S74.6 Brazil: outpatient records and the school census

Brazil's outpatient records (SIA, incident autism diagnoses) enter version 4 under amendment AM-1, which accepted the pilot gate's fixed patient-hash alphabet provisionally, pending the author's veto; the records are separable in one switch. Each SIA fit includes the state-year share of records with a blank diagnosis as a covariate. In the headline, 21,875 of Brazil's 32,869 comparisons are SIA comparisons, and Brazil carries 5.4 percent of the weight. Within Brazil, sampling only, the SIA comparisons give 0.985 (0.966-1.004) and the other Brazilian comparisons 1.024 (0.988-1.060). Dropping SIA gives 0.974 (0.856-1.109). Restricting SIA to state-years with few blank diagnoses leaves only 13 comparisons in 1 municipalities, which is uninformative; coding treated rather than incident cases in those state-years (1,786 comparisons, mostly 2008-2014) gives 0.975 (0.856-1.109).

Brazil's school census (INEP) is report-only. As built, it failed the planted-effect gate in version 3. A specified remodel gave 1.340 (inner 0.984-1.826) but recovered a planted 1.25 with an error of -12.6 percent and returned a future-exposure placebo (theta 0.995 on the log scale) larger than the estimate itself; a within-census contrast against sensory and physical disability gave 1.425 and also failed. The autism share was already rising faster in high-uptake municipalities before their cohorts were exposed.

S74.7 United States: measured sub-state coverage

In the headline, 3,504 of the 4,080 U.S. comparisons use measured sub-state coverage paths: 2,293 infant and 627 child-care paths. Within the United States, sampling only, the sub-state comparisons give 0.939 (0.842-1.047) and the state-series comparisons 0.883 (0.849-0.920). The United States carries 5.2 percent of the cross-country weight, so the between-country variance, not these within-country intervals, governs its contribution.

S74.8 Shared bias and weighting

The placebo-calibrated interval treats bias common to all countries as nearly zero, because the placebo dates estimate the shared component at 0.00 with an upper end of 0.08. Country-level bias, estimated at 0.260, averages down across countries roughly as 0.33 divided by the square root of the effective number of countries (0.33/5.10, about 0.065). A bias shared by every country cannot average down (Figure S120). Table S74-C recomputes the headline interval with an added shared-bias variance; it is computed from the pooled headline row, not approximated.

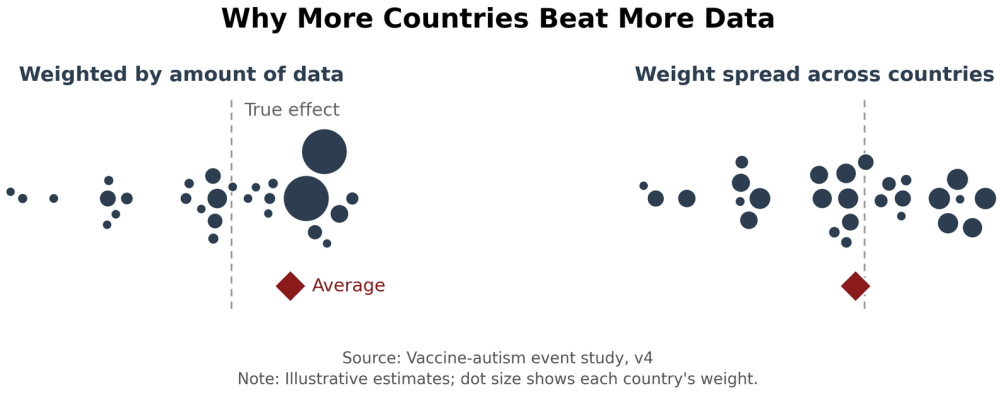


Figure S120. Why spreading the weight across countries narrows the interval. Each country's estimate carries its own bias; weighting by amount of data lets two countries dominate, whereas random-effects weights average country-level biases across many countries. Illustrative estimates; dot sizes are the real version 3 and version 4 weight shares. A bias shared by every country is not averaged away (Table S74-C).

Table S74-C. Headline per-child relative risk 0.973 under an added shared-bias standard deviation s (log scale). Left column: placebo-calibrated variance plus s squared; right column: random-effects variance plus s squared.

Shared bias s	Basis	Calibrated variance + s^2	Random-effects variance + s^2
0.00	v4 placebo estimate	0.855-1.107	0.836-1.131
0.08	upper end of the v4 placebo interval	0.781-1.210	0.771-1.227
0.15	assumed shared allowance	0.705-1.341	0.699-1.354

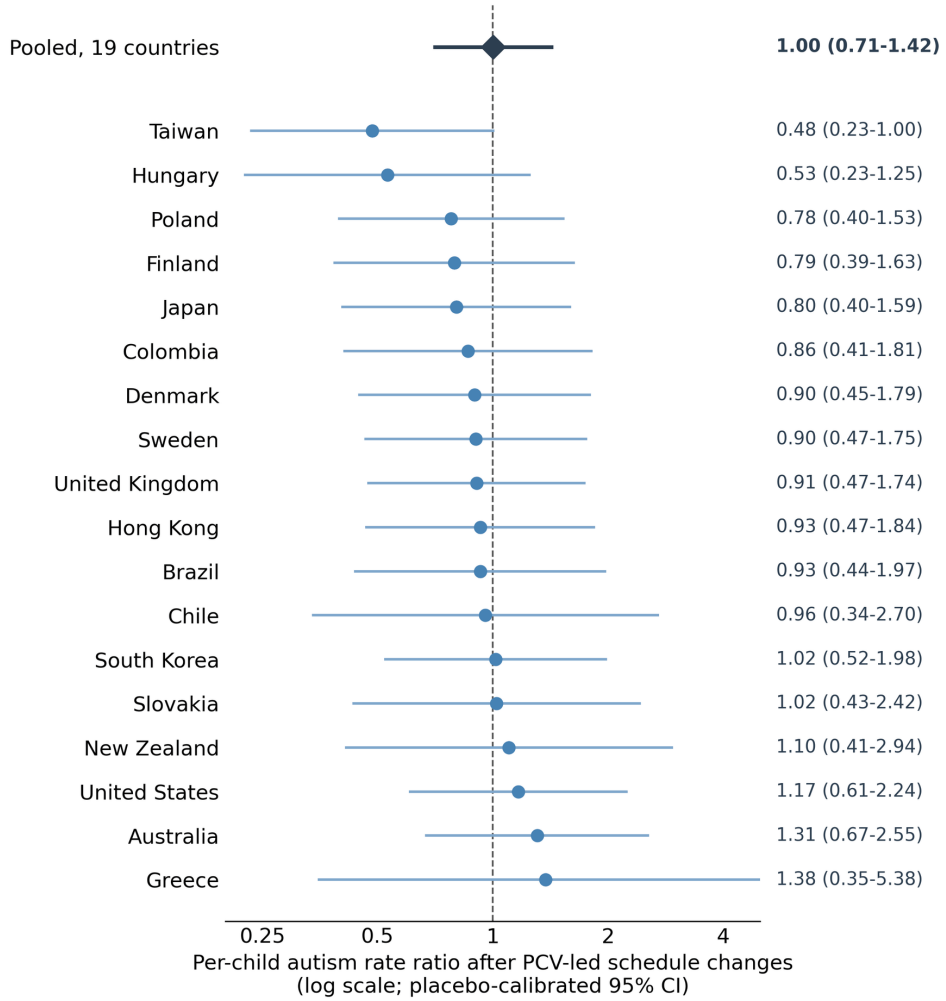
Excluding a 25 percent increase therefore relies on the placebo evidence that no large shared bias exists. Most of the gain in precision over version 3 comes from random-effects weighting rather than new data: with version 3's fixed-information weighting, the version 4 data give a calibrated interval of 0.739-1.265.

S74.9 Ramp test

A step is not the only form harm could take. The ramp model lets the per-child effect grow linearly over the cohorts after the change, with its own placebo variance components, here fitted on the headline set of linear-trend comparisons (35,365 comparisons in 39 countries enter the ramp pool). The post-change slope is -1.73 percent per cohort (placebo-calibrated -8.05 to +5.02). A planted ramp reaching 25 percent by the sixth cohort is recovered with an error of +0.5 percent, but the calibrated interval does not exclude such a ramp.

S74.10 PCV-specific estimate

Restricting the comparisons to changes led by pneumococcal conjugate vaccine gives 1.00 (placebo-calibrated 0.71-1.42) across 1,482 comparisons in 19 countries (Figure S121). This estimate uses the version 3 pooling with fixed-information weights (effective K 3.46), not the version 4 random-effects estimator, and Brazil's school census is excluded. Slovenia's interval is too wide to draw.



PCV-led stacks only; v3 pooling (fixed information weights, Brazil INEP excluded), linear joint trend, window -6 to +6.
Not the v4 random-effects estimator. Omitted from the axis: Slovenia (interval off scale).

Figure S121. Per-child relative risks of autism after schedule changes led by pneumococcal conjugate vaccine, pooled and by country, with placebo-calibrated 95 percent intervals. These use the version 3 pooling with fixed-information weights (Brazil's school census excluded), not the version 4 random-effects estimator; Slovenia's interval is too wide to draw.

S74.11 Data routes not used

The scouting reports for this version list registry and claims data that would add the most information but require the author's application, accreditation, agreement or fee. The largest are German statutory health-insurance claims through the Research Data Centre, Japanese national claims through third-party provision, Nordic registry tables (Danish, Norwegian and Finnish custom extracts), student-level Brazilian school-census records, Spanish ministry tables by province and grade, and Australian disability-scheme tabulations by single year of age. None was requested, so none contributes; the request texts are in the project archive.

S75. Childhood cancer

S75.1 Rule, outcomes and gates

The childhood-cancer outcome family follows a rule hashed before any fit (rule 9000), eleven dated amendments written before the fits they govern, a design file for the later data feeds and a synthesis rule. Outcomes are nested and never combined in one likelihood: all malignant neoplasms, leukaemia, and lymphoid leukaemia (the proxy for acute lymphoblastic leukaemia). Each context is a vaccine introduction or schedule change with a birth-cohort exposure path, fitted on the individual scale with the project's cohort-mixture likelihood. Blocks that fail a source, first-stage, placebo, pretrend, injection or reporting-break gate are report-only. The three Brazilian record systems count the same children's cancers and form one context per outcome. Headline bias allowances follow the data type: context standard deviation 0.20 for registry incidence, 0.35 for deaths and 0.70 for care counts, with a shared component of 0.10.

A positive control had to be recovered before any null was read as informative. HPV programmes recovered the expected reduction in cervical carcinoma in situ in Czechia, 0.42 (95 percent credible interval 0.18-0.83), and in carcinoma in situ with cervical cancer in two Japanese registries, 0.56 (0.33-0.86) and 0.50 (0.30-0.77). Other positive-control fits, including Scotland and Wales, did not recover.

S75.2 Results

Table S75-A. Pooled individual-scale rate ratios (posterior median, 95 percent credible interval) and the posterior probability that the effect exceeds 1.25.

Outcome	Contexts	Headline allowances	P(r > 1.25)	Negative-control calibrated	Doubled allowances
All malignant neoplasms	13	1.00 (0.78-1.28)	0.040	0.97 (0.83-1.14)	0.99 (0.63-1.55)
Leukaemia	10	1.07 (0.79-1.43)	0.142	1.08 (0.86-1.36)	1.06 (0.65-1.72)
Lymphoid leukaemia	9	1.16 (0.83-1.62)	0.339	1.18 (0.89-1.56)	1.16 (0.68-1.95)

The calibrated variant for all malignant neoplasms excludes 1.25; no other row does. Power is limited: in whole-pipeline simulations with no effect, the analysis excludes 1.25 with probability 0.43 for all malignant neoplasms, 0.29 for leukaemia and 0.24 for lymphoid leukaemia, and it detects a true 1.5 with probability 0.86, 0.72 and 0.62. Lymphoid leukaemia leans above one; its interval is wide and its context estimates are mixed in sign.

On the observed scale, the summative event study of the same series gives pooled steps of 1.010 (sampling-only 0.990-1.041) for all malignant neoplasms across 13 contexts and 86 series, and 1.010 (0.970-1.051) for leukaemia across 10 contexts and 36 series (Figure S122).

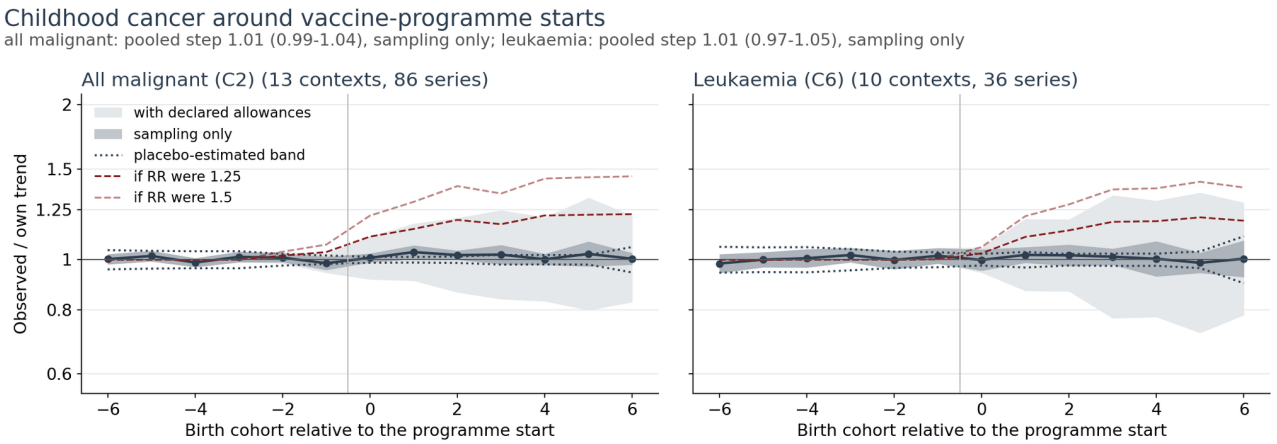


Figure S122. Childhood cancer around vaccine-programme starts: observed counts relative to each series' own trend, pooled across contexts for all malignant neoplasms (left) and leukaemia (right). Dark bands are sampling-only, light bands add the assumed allowances, dotted lines are placebo-estimated bands, and dashed lines show the path a rate ratio of 1.25 or 1.5 would produce. Observed-scale steps, not individual-scale rate ratios.

S76. Additional design-specific evidence

These passages appeared in the main text of the journal article until the length revision of October 2026 (author decision 4); they are reproduced verbatim, and the journal keeps a one-sentence pointer at each place.

S76.1 Negative controls, context by context (journal Section 4.4)

In the United States, the core PCV context leans harmful on autism, 1.17 (0.97-1.39) with P(r > 1) of 0.948. The same joint likelihood, with autism replaced by other special-education categories treated as candidate controls, leans as far or further: traumatic brain injury 1.65 (0.77-4.29), orthopedic impairment 1.43 (0.91-2.26), hearing impairment 1.41 (1.05-1.91) and visual impairment 1.45 (0.74-2.72). Under the stated rule only hearing impairment is precise enough to count, and it leans like autism, so the verdict is MIXED: the rule establishes neither that the autism lean is shared drift nor that it is specific to autism. The same controls carry rotavirus-2006 leans of their own (traumatic brain injury 2.37, 1.09-4.58; hearing 1.41, 1.07-1.82) where autism gives 0.95. A separate outcome system shows the bias directly. In the NSCH survey dose-response design, parent-reported brain injury or concussion rises 1.12 (1.02-1.23) per ten points of PCV coverage, a bias signal in the harm direction, against autism 1.10 (0.93-1.32) in the same children. Child deaths from unintentional injuries on state-by-birth-cohort cells give a PCV "effect" of 2.59 (1.15-5.76). The rule classes it UNINFORMATIVE, because its interval is 4.4 times as wide as autism's, and on the same cells the future-exposure placebo returns an "effect" in the opposite direction (0.18, 0.10-0.39), so the machinery finds associations there that are not tied to the vaccine's timing. Autism relative to injury deaths on the core's own state-cohort units is 1.03 (0.69-1.52).

In Brazil, the maternal DTaP design gives autism 0.77 (0.51-1.14) on 108 Census cells, and all-injury hospital admissions on the same cells, with the same exposure path, give 0.51 (0.29-0.83): a bias signal in the protective direction, and the stated reading is DESIGN BIAS DETECTED. Falls give 0.38 (0.15-0.86), and the future-exposure placebo fails for all injuries (0.63, 0.43-0.90) while passing for autism. The same Census design detects harm-direction bias for hepatitis A, where autism gives 2.17 (1.36-3.98) and all injuries 2.01 (1.15-4.23), and protective-direction bias for PCV10 with MenC (all injuries 0.44, 0.28-0.75). Deaths from unintentional injuries on the same 108 cells, from the mortality register rather than hospital records, give 0.36 (0.15-0.79), a second protective-direction bias signal; autism relative to those deaths is 1.92 (0.97-3.57).

In Hong Kong, the PCV design gives autism 0.89 (0.66-1.16) on 60 grade-by-year cells, but physical disability 1.78 (1.23-2.90) and physical plus visual disability 1.50 (1.09-2.22); the system reading of its rule is BIAS SIGNAL. That rule was written with the deciding estimate already known, and simulations with cohort-level noise of the observed size show that the design's nominal intervals under-cover and put the physical-disability value at a 12-15 percent two-sided tail, so the reading is a disclosure of possible design bias rather than a calibrated test; the composite of physical, visual and hearing categories, whose class was unknown when the rule was written, is 1.15 (0.90-1.47), CONSISTENT. The physical-disability interval excludes one even at a Bonferroni 99 percent level (1.06-3.47), although that control's two-sided placebo family fails in 3 of 20 member-by-setting intervals. Police-reported road-traffic casualties give 0.66

(0.48-0.92), a bias signal in the protective direction that depends on the window (0.98, 0.59-1.63, on 2013-2021), so the Hong Kong design can return spurious "effects" in both directions. Injury deaths cannot test it (0.89, 0.26-3.74).

Elsewhere the controls are clean at their precision. England supplies the cleanest test. Police-recorded road casualties among children on the cells of England's rotavirus design (112 of its 128 age-by-year cells, the census years 2011-2024) give 1.00 (0.87-1.16), beside autism of 0.97 (0.89-1.06) refitted on the same 112 cells, and injury admissions give 0.99 (0.57-1.57) by diagnosis and 0.98 (0.53-1.60) by external cause. None of the fourteen primary UK controls is a bias signal; one of 53 secondary components, poisoning admissions (0.41, 0.19-0.85), signals in the protective direction. Swedish injury diagnoses in inpatient and outpatient care on the PCV design's own 756 cells give 1.06 (0.98-1.13), CONSISTENT and nearly as precise as England's control, beside autism of 1.01 (0.79-1.56). Northern Ireland's police-recorded casualties (1.02, 0.37-2.04), Canada's asthma incidence (0.71, 0.48-1.06), Tanzania's pupils with albinism (1.00, 0.69-1.28), Japan's unintentional injury deaths (0.86, 0.46-1.61) and Colombia's hearing, motor and visual disability categories are CONSISTENT. Colombia's non-fatal transport injuries give 1.36 (0.99-1.93), CONSISTENT by the rule but at its edge, and dropping the 2020-2021 school years, when lockdowns cut forensic examinations, moves Colombian autism from 0.76 (0.51-1.07) on all cells to 1.31 (0.89-1.98) on the control's 335 cells: the Colombian estimate depends on the lockdown years. German PCV injury deaths, 1.01 (0.50-1.81), read CONSISTENT only through a grouped-age implementation note added after the rule; the German and Czech care controls and most UK injury-death controls are too imprecise to read, and Quebec has no control on identical cells. Consistency is not validation. In Chile, a report-only design, an injected doubling would still have been classed CONSISTENT, and in Japan the injury estimate sits where autism's does, so the control neither supports nor undercuts the Japanese estimate.

S76.2 Type 1 diabetes: allocated comparisons and the project's own analyses (journal Section 6.2)

Three comparisons allocated vaccination by randomization or birth date; none supports a precision or exclusion claim. The Finnish Hib trial contrasted four early doses with one dose at 24 months (Eskola et al. 1990), and ten-year diabetes risk did not differ significantly (RR 1.06; Karvonen, Cepaitis, and Tuomilehto 1999). A reanalysis against a pre-trial cohort reported relative risk 1.26 with clusters located by searching the follow-up data (Classen and Classen 2002); as a non-randomized comparison across birth years it receives zero weight. A follow-up of the Northern California HbOC trial cohort reported no association (Black et al. 2002). In the randomized Finnish rotavirus trial extension only 30 percent of participants returned a questionnaire, and unequal respondent arms mean that neither the diabetes counts nor the celiac comparison inherits randomization (Hemming-Harlo et al. 2019; Supplement S49.1).

The project's own diabetes analyses are weak. EURODIAB introduction-date models on 124 centre-period rates (Patterson et al. 2019) give Hib and PCV ratios below one without pretrend tests, and single-country deletions spread an earlier PCV estimate from 0.687 to 1.137; these inverse ratios are not evidence of protection. Age-gradient models change sign when centre trends are added, and their two harmful-direction Hib estimates, 1.274 (1.013-1.602) and 1.276 (1.009-1.614), do not survive Holm adjustment (Supplement S47). Swedish annual contrasts (Rangert et al. 2023) give 1.074 (0.793-1.455), with intervals that undercover and low power against a ten percent step (Supplement S50). Finland's published decline, 0.79 (0.71-0.86), is matched by an earlier pre-vaccine cohort (Parviainen et al. 2024), and an eight-country compilation re-reports earlier studies (Rogers 2025; Supplement S51). Person-level rotavirus estimates are not uniformly null: 1.03, with a partial-series HR of 1.50 (0.81-2.77), in the Vaccine Safety Datalink (Glanz et al. 2020); 1.09 in MarketScan (Burke et al. 2020); and 0.67 (0.54-0.83) in Optum (Rogers, Basu, and Kim 2019). Each needs appraisal, not a verdict by direction (Supplement S49). A dependence-sensitive synthesis of adjusted rotavirus receipt associations from Norway, the United Kingdom and three U.S. data systems gives a type 1 diabetes HR of 0.940 (95 percent credible interval 0.688-1.301). Selecting one U.S. system at a time and retaining full-series, any-dose and risk-ratio contrasts separately avoids treating overlapping insurance populations or alternative exposure definitions as independent. The coeliac-disease HR family gives 1.032 (0.750-1.417). These observational summaries do not enter the policy-event autism hierarchy and do not establish protection or rule out moderate chronic effects (Supplement S69).

S76.3 Aluminum, asthma and youth obesity (journal Section 6.3)

Andersson et al. (2025) related cumulative vaccine aluminum by age two to 50 outcomes among Danish children. Per-milligram HRs had point estimates below one for any neurodevelopmental disorder, ADHD, autoimmune and atopic or allergic disorders (Supplement S49.1); atopic dermatitis, 1.02 (1.00-1.04), is the only main-text outcome whose lower limit is 1.00. A correction replaced the supplementary material (Annals of Internal Medicine 2025). The corrected Asperger estimate among 2007-2018 births is 1.67 (1.01-2.77) on 51 events without multiplicity correction, against 0.75 (0.41-1.40) originally, and Asperger syndrome at age eight is 1.02 (0.93-1.12). A Vaccine Safety Datalink cohort of 326,991 children associated aluminum before age two with persistent asthma, HR per milligram 1.19 (1.14-1.25) without eczema and 1.26 (1.07-1.49) with it (Daley et al. 2023), against a Danish 0.96 (0.94-0.98). That association, which its authors note may reflect residual confounding, receives zero weight but is not dismissed because the Danish estimate is inverse: for chronic outcomes other than autism, the observational aluminum evidence is mixed. For epilepsy, all person-level evidence is observational. A Vaccine Safety Datalink case-control study reports that no adjusted odds ratio in its primary analysis exceeded one (McClure et al. 2026), but without accessible intervals it supplies no upper bound on harm; cohort HRs are 1.01 after DTaP-IPV-Hib (Sun et al. 2012) and 1.07 after pandemic influenza vaccination (Håberg et al. 2018), both with intervals including one (Supplement S49.1).

For youth obesity, modeled from published YRBS state-year-grade prevalences, the informative result is a failure. Obesity is 0.970 (0.954-0.985) per standard deviation of PCV7 rollout intensity, but every pretrend and 1993 placebo test rejects, and the fictitious 1993 step, 0.967 (0.953-0.982), is as large as the main estimate. This is not evidence that PCV7 prevents obesity; it shows that the rollout index tracks pre-existing state-cohort drift, which Section 4.7 uses to falsify the autism coefficient. The RotaShield window gives 1.013 (0.977-1.051) and ever-diagnosed asthma 0.988 (0.959-1.019), with timing diagnostics that do not reject but have unassessed power, and WIC toddler obesity 1.013-1.015, with timing checks withheld for too few pre-event cohorts (Supplement S48). Because YRBS samples only students attending school, it cannot bound effects on youths whose impairments reduce attendance. A Swedish pertussis-component contrast gave dispensed asthma medication at age 15 an adjusted OR of 0.99 (0.95-1.03), without timing diagnostics (Vogt et al. 2014).

S76.4 Disability and psychiatric panels (journal Section 6.4)

Hong Kong special-education screens give false-discovery-rate survivors per ten points of PCV coverage for intellectual disability, 1.033 (1.009-1.057), and physical disability, 1.049 (1.016-1.084); the physical-disability "effect" is a bias signal (Section 4.4). In the U.S. core cells, other health impairment is 1.40 (1.12-1.77) and intellectual disability 1.18 (0.94-1.49), but the future-exposure placebo fails for both. The SchoolVaxView other-health-impairment share rises 1.168 (1.013-1.347) per ten points of MMR coverage without multiplicity adjustment. Peruvian hearing, vision and autism registrations rise with MMR's service share, as do their future-exposure placebos. Brazilian maternal Tdap gives any disability 1.019 (1.002-1.035) unweighted, nonsignificant when weighted. New Jersey's school-compliance panel has two nominally positive coefficients, for multiple disabilities and orthopedic impairment, without stored placebos. Brazilian hospitalization families yield no discovery among 176 estimates (one nominal estimate above one, outside the primary specification) and their positive controls fail; Taiwan's 288 tests yield none.

Broader disability panels relate area coverage to non-autism outcomes. None has a validated policy first stage; their harm-direction rows are listed with their diagnostics in Supplement S48.1. Norwegian primary-care claims yield no false-discovery-rate survivor among 88 estimates, but six of eight nominal estimates are harmful-direction, among them measles coverage with anxiety or depression symptoms, 1.131 (1.018-1.255), and with musculoskeletal diagnoses; their matched future placebos do not reject, and the panel is exploratory only. All eight nominally positive rows of the stratified French file are intellectual disability at ages 5-9, each passing its matched placebo, for example MMR1 1.347 (1.056-1.719); none survives adjustment (smallest q 0.085), and at ages 10-14 five of six intellectual-disability placebos are nominally positive. Three of six pooled-cohort intellectual-disability MMR rows are also nominally positive, but in the same claims impossible future-cohort coverage predicts childhood psychiatric outcomes (DTP3 placebo 0.827, 0.726-0.941) and the MMR2 future placebo equals its estimate (1.100).

S76.5 Convergence on shared cells (journal Section 4.7)

Designs that reuse the same cells cannot replicate one another: their agreement shows robustness to estimator and functional form, and it cannot rule out a bias they share (Figure 4). Where designs share cells, they agree. England rotavirus gives 0.94 (0.86-1.02) on its own data and 0.95 (0.82-1.10) as a jump; Northern Ireland gives 1.01 (0.88-1.16) and 0.91 (0.73-1.15); Sweden PCV gives 1.01 (0.79-1.56) and 0.98 (0.77-1.26) in the all-ages block. U.S. hepatitis A 1999 leans above one in the jump (1.12, 0.69-1.82) and in two synthetic controls (1.18 and 1.23), each interval including one. Where designs disagree, the disagreement is diagnostic. Canadian synthetic controls sit at 1.76-4.99 with every lower bound at zero, because only ages one to three are exposed in the clean window and the conversion divides a noisy gap by a small first stage; the hierarchy gives 0.99 (0.32-1.89) on the same cells. U.S. rotavirus 2006 gives 0.95 (0.76-1.19) in the core and 1.31 (0.74-2.33) as a jump on the same cells, and its negative controls carry leans of their own. Of the 39 synthetic-control rows, 15 are discarded and none enters the pool.

S76.6 Cohort studies of MMR (journal Section 2.1)

For MMR, Madsen et al. (2002) reported no elevated adjusted association among 537,303 Danish children, and Hviid et al. (2019) reported none overall, no post-vaccination clustering and no excess in familial-risk strata among 657,461 children. Jain et al. (2015) reported no harmful association among 95,727 U.S. children, including those with an older autistic sibling; Smeeth et al. (2004) and Uno et al. (2015) add null-centered calibration. A 2026 U.S. EHR cohort of 2,560,035 children reported an adjusted HR of 0.97 (99 percent CI 0.91-1.03) for first MMR before age two, with a near-null negative-control exposure (Burstain, Jacques, and Burstain 2026). Farrington, Miller, and Taylor (2001) and Honda, Shimizu, and Rutter (2005) provide cohort-timing and withdrawal histories; the directionally informative Yokohama falsification in this paper comes from an explicit coverage path and placebo structure.

S76.7 The most precise single jumps (journal Section 4.1)

The most precise estimates come from large national introductions. England's 2006 PCV introduction, a step of about 89 percentage points, gives 0.941 (0.862-1.027), but it measures the whole September 2006 booster package, with which PCV is collinear (a Hib/MenC booster began at the same time); it is outside the pool because it shares the cells of the pooled England rotavirus block; and the same design returns a dip as large for speech, language and communication needs, so the dip is not autism-specific. England's 2013 rotavirus introduction gives 0.951 (0.825-1.097); no specification passes its placebo check, so the widest calibrated interval is reported. U.S. PCV7 gives 1.042 (0.883-1.230) in the IDEA state panel and 1.027 (0.569-1.853) in California. The other 27 calibrated estimates, in England, Northern Ireland, Scotland, the United States, Sweden, Germany and France (France's 2018 extension of compulsory vaccination being a jump of 18.8 points on a psychiatric outcome that is not autism-specific), run from 0.65 to 1.93, every interval including one; six uncalibrated estimates, among them the Yokohama fit, which has no finite calibrated interval, and a California DDS design the checker judged invalid, are listed with them (Supplement S60.3, Table S60-A). Those with harm-direction specifications return in Section 5.1.

S76.8 Colombia's school census (journal Section 4.3)

One input moves the synthesis materially. Colombia's school census includes the 2020-2021 school years, when closures disrupted school reporting and cut the forensic injury control (Section 4.4). Refitting Colombia without those years moves its own estimate from 0.76 (0.51-1.07) to 1.32 (0.90-1.99), the shared mean to 1.045 (0.654-1.655) and the common effect to 1.054 (0.697-1.591) (Supplement S62.8). A common effect of 1.5 is then not excluded even under the assumed allowances, and a common effect of 1.10 becomes ordinary (the lockdown-free shared mean lies at the 30th percentile of replicates generated under it), but a common effect of 1.25 still produces an estimate that low in only 0.25 percent of replicates. The synthesis keeps the full series that passed the gates. All these percentiles assume no bias; the negative controls probe, but cannot establish, that assumption (Section 4.4).

S76.9 Unpooled block posteriors (journal Section 5.1)

The nine unpooled block posteriors qualify by interval or by posterior probability. A Brazilian hepatitis A census variant gives autism 2.17 (1.36-3.98), but its future-exposure placebo is 1.85 (1.23-3.09) and the same design returns an all-injury bias signal (Section 4.4). French hepatitis B coverage gives other neurologic conditions 1.46 (1.17-1.81) with a placebo of 1.37 (1.04-1.76), and chronic respiratory conditions 1.15 (0.94-1.35) with a placebo of 1.58 (1.28-1.95). The remaining six, Polish, Swedish, Taiwanese and Welsh blocks, have intervals including one (Supplement S59.5).

S76.10 Origins of the claim (journal Section 1)

The vaccine-autism claim belongs to an older movement combining objections to coercion with allegations that vaccines are unnatural, toxic or protected by interested authorities (Wolfe and Sharp 2002; Poland and Jacobson 2011; Kata 2012; Dube, Vivion, and MacDonald 2015). The scare accelerated after a 1998 Lancet case series of twelve children, retracted in 2010. Deer (2011) documented altered clinical histories and an undisclosed litigation-related conflict; the BMJ's editors judged it fraudulent (Editors of The Lancet 2010; Godlee, Smith, and Marcovitch 2011). Allegations shifted from MMR to thimerosal, antigens, aluminum, prenatal exposure, simultaneous administration, "too many too soon" and susceptible subgroups (Offit et al. 2002; Poland and Jacobson 2012). Each shift changes the test; none makes the claim unfalsifiable.

S76.11 Mechanisms and their predicted signatures (journal Section 2.2)

Mechanisms predict persistence, dose response, timing or susceptibility; plausibility cannot veto data. The measles-enteropathy hypothesis predicted persistent vaccine-strain virus and ordered gastrointestinal and developmental onset; a blinded three-laboratory study found neither (Hornig et al. 2008). Prenatal trajectories, cortical organization, heritability and neuronal gene networks inform autism biology without identifying vaccination as its initiating exposure (Stoner et al. 2014; Hazlett et al. 2017; Bai et al. 2019; Satterstrom et al. 2020). Ingredient and schedule theories predict dose response, stacking, thresholds and critical windows. The thimerosal, antigen and aluminum studies above show no robust autism gradient. Prenatal exposure requires pregnancy-specific data, not infant coverage; Sections 4.6 and 5.2 test the population analogues available here.

S76.12 What the negative controls imply (journal Section 4.4)

Two implications follow. Agreement among designs on the same cells cannot establish specificity where the controls lean too (Section 4.7). And the bias the controls reveal runs in both directions, which the hierarchy represents through its context-bias terms: Hong Kong's physical-disability lean, 0.58 on the log scale, is smaller than the 0.70 bias standard deviation the model assigns to that context. A calibration of the synthesis against these controls is reported in Section 4.3. Counting both rounds of controls, each source's primary reading rows show a bias signal in four of the sixteen hierarchy contexts (the U.S. PCV and rotavirus-2006 contexts, Hong Kong and Brazil; seven when any single category counts, the extra three being single fracture, poisoning and composite-injury codes), only consistent controls in nine and only uninformative ones in two, and Quebec has only an analogue control; of all 452 fitted negative-control estimates, 28 are bias signals, 195 are consistent and 229 are uninformative. About eleven of the 223 informative estimates would exclude one by chance at a two-sided 5 percent rate; 21 of the 28 that do come from the U.S., Brazilian and Hong Kong designs, and most of the rest are single diagnosis codes.

S76.13 Conditions that are not negative controls (journal Section 5.2)

Other conditions are not automatically negative controls. Intellectual disability, speech and language, ADHD, epilepsy, psychiatric outcomes, autoimmune disease and allergy are claim-relevant whenever a broad injury theory includes them, and Section 6 reports them; sensory and physical categories serve as controls under stated rules for products with no proposed mechanism for them. The preferred autism-plus-intellectual-disability-plus-speech pool is RR 1.000 (0.964-1.036), arguing against relabeling as the reason the focal coefficient is near one. Positive controls behave as expected where they can be run: in Brazil the 2007-versus-2005 infant/adult gastroenteritis contrast is 0.701, consistent in direction with infection prevention, although dependence and weak pretrend diagnostics limit its interpretation.

S76.14 Chronic-harm claims from utilization contrasts (journal Section 6.1)

No design here estimates the contrast behind headline chronic-harm claims. In Florida Medicaid claims for 47,155 nine-year-olds, children with billed vaccination visits had any neurodevelopmental disorder at OR 3.12 (2.85-3.41) (Mawson and Jacob 2025), and an unpublished Henry Ford Health draft reports an adjusted hazard ratio of 2.54 (2.16-2.97) for chronic conditions among 16,511 vaccinated versus 1,957 unvaccinated children while noting that unvaccinated children use less healthcare (Lamerato et al. 2025). Such contrasts are open to healthcare-utilization and ascertainment bias and receive zero weight.

S77. Matched contrasts, source coverage and uptake-path sensitivity

S77.1 The specified control list

Controls were specified from record structure alone before any v5 fit (RULE_9160 N): 158 rows in the factory's specified list, of which 77 in 31 sources are the co-primary controls, 57 form the sensitivity set (42 of them hearing-related, because meningitis and congenital rubella are vaccine-preventable causes of deafness) and 24 are never used. No co-primary control falls under the fixed exclusions (intellectual or learning disability; speech, language and communication; social-emotional, behavioural and attention disorders; specific learning difficulties; developmental delay; multiple, other, unassessed and pooled categories; all-SEN and all-discharge totals; high ability). Where a source carries a control in several bases, only the basis of its primary autism code is used (U.S. IDEA all-ages codes; Greece's Ministry-of-Education basis; Slovakia's integrated placement; both Japanese bases, as for autism).

<!-- S77-A -->

Table S77-A. Specified control categories by source and role (from DECLARED_NC_LIST.csv and its rest-of-world additions, DECLARED_NC_LIST_ROW_ADDITIONS.csv). Vaccine pathway is the specified plausibility of a vaccine effect on the category.

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
AUS	AUS_NDIS_PLAN_BUDGETS	AUS_NDIS_CEREBRAL_PALSY	registry	primary control	minor	v5 specified list
AUS	AUS_NDIS_PLAN_BUDGETS	AUS_NDIS_SPINAL_CORD_INJURY	registry	primary control	none	v5 specified list
AUS	AUS_NDIS_PLAN_BUDGETS	AUS_NDIS_VISUAL	registry	primary control	minor	v5 specified list
AUS	AUS_NDIS_PLAN_BUDGETS	AUS_NDIS_HEARING	registry	sensitivity set	plausible	v5 specified list
BRA	BRA_INEP_SINOPSE_TEA	SENSORY_PHYSICAL	school_census	sensitivity set	plausible	v5 specified list
BRA	BRA_SIA_F84	NC_ST_TREATED	claims	primary control	none	v5 specified list
CAN	(not written)	SPED_C	school_census	never used	excluded	v5 specified list
CAN	CAN_BC_SD_NC	SPED_A	school_census	primary control	minor	v5 specified list
CAN	CAN_BC_SD_NC	SPED_E	school_census	primary control	minor	v5 specified list
CAN	CAN_BC_SD_NC	SPED_B	school_census	sensitivity set	plausible	v5 specified list
CAN	CAN_BC_SD_NC	SPED_D	school_census	sensitivity set	plausible	v5 specified list
CAN	CAN_BC_SD_NC	SPED_F	school_census	sensitivity set	plausible	v5 specified list
CAN	CAN_CCDSS_NC	ASTHMA	claims	sensitivity set	plausible	v5 specified list
CHL	CHL_DEIS_NC	APPENDICITIS_K35_K38	hospital_discharge	primary control	none	v5 specified list
CHL	CHL_DEIS_NC	INJURY_EXCL_T78_T80_T88	hospital_discharge	primary control	none	v5 specified list
CZE	CZE_NRHZS_NC_DISTRICT	APPENDICITIS_K35_K38	claims	primary control	none	v5 specified list
CZE	CZE_NRHZS_NC_DISTRICT	CONGENITAL_Q_PUBLISHED	claims	primary control	none	v5 specified list
CZE	CZE_NRHZS_NC_DISTRICT	CP_G80	claims	primary control	minor	v5 specified list
CZE	CZE_NRHZS_NC_DISTRICT	INJURY_EXCL_T78_T80_T88	claims	primary control	none	v5 specified list
CZE	CZE_NRHZS_NC_DISTRICT	HEARING_H90_H91	claims	sensitivity set	plausible	v5 specified list
DEU	DEU_DESTATIS_INJURY_LAND	INJ_EXCL_T78_T80_T88	hospital_discharge	primary control	none	v5 specified list
DNK	DNK_IND04_HOSP_PATIENTS	DNK_IND04_B35	hospital_discharge	primary control	none	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	FISICA	school_census	primary control	minor	v5 specified list

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
ECU	ECU_MINEDUC_DISABILITY_GRADE	VISUAL	school_census	primary control	minor	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	AUDITIVA	school_census	sensitivity set	plausible	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	SORDOCEGUERA	school_census	sensitivity set	plausible	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	DISC_ANY	school_census	never used	excluded	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	INTELECTUAL	school_census	never used	excluded	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	LENGUAJE	school_census	never used	excluded	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	NO_DECLARA	school_census	never used	excluded	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	PSICOSOCIAL	school_census	never used	excluded	v5 specified list
ECU	ECU_MINEDUC_DISABILITY_GRADE	RETOS_MULTIPLES	school_census	never used	excluded	v5 specified list
ESP	ESP_MEFPD_NEE_TYPES_PROV	MOTOR	school_census	primary control	minor	v5 specified list
ESP	ESP_MEFPD_NEE_TYPES_PROV	VISUAL	school_census	primary control	minor	v5 specified list
ESP	ESP_MEFPD_NEE_TYPES_PROV	HEARING	school_census	sensitivity set	plausible	v5 specified list
FRA	FRA_RHEOP_AGE7	MOTOR	registry	primary control	minor	v5 specified list
FRA	FRA_RHEOP_AGE7	T21_IQ50	registry	primary control	none	v5 specified list
FRA	FRA_RHEOP_AGE7	VISUAL	registry	primary control	minor	v5 specified list
FRA	FRA_RHEOP_AGE7	HEARING	registry	sensitivity set	plausible	v5 specified list
FRA	FRA_RHEOP_AGE7	SENSORY	registry	sensitivity set	plausible	v5 specified list
FRA	FRA_RHEOP_AGE7	INTELLECTUAL	registry	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	DS	school_census	primary control	none	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	PD	school_census	primary control	minor	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	VI	school_census	primary control	minor	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	HI	school_census	sensitivity set	plausible	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	MSI	school_census	sensitivity set	plausible	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	MLD	school_census	never used	excluded	v5 specified list

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	NSA	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	OTH	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	PMLD	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	SEMH	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	SLCN	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	SLD	school_census	never used	excluded	v5 specified list
GBR	GBR_ENG_SEN_OTHER_NEED_UTLA	SPLD	school_census	never used	excluded	v5 specified list
GBR	GBR_SCT_ASN_OTHER_COUNCIL	PHYS_MOTOR	school_census	primary control	minor	v5 specified list
GBR	GBR_SCT_ASN_OTHER_COUNCIL	VI	school_census	primary control	minor	v5 specified list
GBR	GBR_SCT_ASN_OTHER_COUNCIL	HI	school_census	sensitivity set	plausible	v5 specified list
GBR	GBR_WAL_ALN_OTHER_LA	VI	school_census	primary control	minor	v5 specified list
GBR	GBR_WAL_ALN_OTHER_LA	HI	school_census	sensitivity set	plausible	v5 specified list
GBR	GBR_WAL_ALN_OTHER_LA	MSI	school_census	sensitivity set	plausible	v5 specified list
GBR	GBR_WAL_ALN_OTHER_LA	PMED	school_census	sensitivity set	plausible	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	MOTOR_ALL	school_census	primary control	minor	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	MOTOR_MOE	school_census	primary control	minor	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	VISION_ALL	school_census	primary control	minor	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	VISION_MOE	school_census	primary control	minor	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	HEARING_ALL	school_census	sensitivity set	plausible	v5 specified list
GRC	GRC_ELSTAT_SPED_OTHER	HEARING_MOE	school_census	sensitivity set	plausible	v5 specified list
HKG	HKG_EDB_SEN	HKG_SEN_PHYSICAL_DISABILITY	school_census	primary control	minor	v5 specified list
HKG	HKG_EDB_SEN	HKG_SEN_VISUAL_IMPAIRMENT	school_census	primary control	minor	v5 specified list
HKG	HKG_EDB_SEN	HKG_SEN_HEARING_IMPAIRMENT	school_census	sensitivity set	plausible	v5 specified list
HUN	HUN_KSH_SEN_NC	PHYSICAL	school_census	primary control	minor	v5 specified list
HUN	HUN_KSH_SEN_NC	VISUAL	school_census	primary control	minor	v5 specified list
HUN	HUN_KSH_SEN_NC	HEARING	school_census	sensitivity set	plausible	v5 specified list
ISR	ISR_CBS_PUB1980_2023	ISR_REG_PHYSICAL	registry	primary control	minor	rest-of-world addition

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
ISR	ISR_CBS_PUB1980_2023	ISR_REG_VISION	registry	primary control	minor	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_CHRONIC	registry	sensitivity set	plausible	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_HEARING	registry	sensitivity set	plausible	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_INTELLECTUAL	registry	never used	excluded	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_OTHER	registry	never used	excluded	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_PSYCHIATRIC	registry	never used	excluded	rest-of-world addition
ISR	ISR_CBS_PUB1980_2023	ISR_REG_UNKNOWN	registry	never used	excluded	rest-of-world addition
JPN	JPN_MEXT_SSC	JPN_MEXT_SSC_PHYSICAL_SENSORY	school_census	sensitivity set	plausible	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_LOW_VISION_FY	school_census	primary control	minor	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_LOW_VISION_MAY1	school_census	primary control	minor	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_PHYSICAL_FY	school_census	primary control	minor	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_PHYSICAL_MAY1	school_census	primary control	minor	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_HEALTH_FY	school_census	sensitivity set	plausible	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_HEALTH_MAY1	school_census	sensitivity set	plausible	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_HEARING_FY	school_census	sensitivity set	plausible	v5 specified list
JPN	JPN_MEXT_TSUKYU	JPN_TSUKYU_HEARING_MAY1	school_census	sensitivity set	plausible	v5 specified list
KOR	KOR_NISE_SPED_NC	KOR_NISE_PHYSICAL	school_census	primary control	minor	v5 specified list
KOR	KOR_NISE_SPED_NC	KOR_NISE_VISUAL	school_census	primary control	minor	v5 specified list
KOR	KOR_NISE_SPED_NC	KOR_NISE_HEALTH	school_census	sensitivity set	plausible	v5 specified list
KOR	KOR_NISE_SPED_NC	KOR_NISE_HEARING	school_census	sensitivity set	plausible	v5 specified list
LTU	LTU_OSP_SEN_PDD	SEN_fiziniai	school_census	primary control	minor	v5 specified list
LTU	LTU_OSP_SEN_PDD	SEN_rega	school_census	primary control	minor	v5 specified list
LTU	LTU_OSP_SEN_PDD	SEN_implantai	school_census	sensitivity set	plausible	v5 specified list
LTU	LTU_OSP_SEN_PDD	SEN_klausa	school_census	sensitivity set	plausible	v5 specified list
LTU	LTU_OSP_SEN_PDD	SEN_kurcnerega	school_census	sensitivity set	plausible	v5 specified list
MEX	MEX_DGIS_SAEH	APPENDICITIS_K35_K38	hospital_discharge	primary control	none	v5 specified list
MEX	MEX_DGIS_SAEH	DOWN_Q90_ANY	hospital_discharge	primary control	none	v5 specified list
MEX	MEX_DGIS_SAEH	INJURY_EXCL_T78_T80_T88	hospital_discharge	primary control	none	v5 specified list
MEX	MEX_DGIS_SAEH	ALL_CHILD	hospital_discharge	never used	excluded	v5 specified list
NOR	NOR_KPR_PRIMARY_CARE	NOR_KPR_K70_K99	claims	primary control	minor	v5 specified list

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
NOR	NOR_KPR_PRIMARY_CARE	NOR_KPR_Skader	claims	primary control	none	v5 specified list
NZL	NZL_MOH_DSS_DEMOGRAPHIC	NZL_DSS_SENSORY_ANY	registry	sensitivity set	plausible	v5 specified list
NZL	NZL_MOH_DSS_DEMOGRAPHIC	NZL_DSS_SENSORY_FY	registry	sensitivity set	plausible	v5 specified list
NZL	NZL_MOH_DSS_DEMOGRAPHIC	NZL_DSS_SENSORY_PRINCIPAL	registry	sensitivity set	plausible	v5 specified list
PER	PER_MINEDU_SIAGIE_NC	PHYSICAL	school_census	primary control	minor	v5 specified list
PER	PER_MINEDU_SIAGIE_NC	VISUAL	school_census	primary control	minor	v5 specified list
PER	PER_MINEDU_SIAGIE_NC	HEARING	school_census	sensitivity set	plausible	v5 specified list
PER	PER_MINEDU_SIAGIE_TEA	SENSORY_PHYSICAL	school_census	sensitivity set	plausible	v5 specified list
POL	POL_NFZ_REFUND_NC	APPENDICITIS_K35_K37	claims	primary control	none	rest-of-world addition
POL	POL_NFZ_REFUND_NC	CEREBRAL_PALSY_G80	claims	primary control	minor	rest-of-world addition
POL	POL_NFZ_REFUND_NC	CONGENITAL_NONCHROM_Q00_Q89	claims	primary control	none	rest-of-world addition
POL	POL_NFZ_REFUND_NC	INJURY_HEAD_S00_S09	claims	primary control	none	rest-of-world addition
POL	POL_NFZ_REFUND_NC	INJURY_OTHER_S10_T14	claims	primary control	none	rest-of-world addition
POL	POL_NFZ_REFUND_NC	VISUAL_IMPAIRMENT_H54	claims	primary control	minor	rest-of-world addition
POL	POL_NFZ_REFUND_NC	CHROMOSOMAL_Q90_Q99	claims	sensitivity set	none	rest-of-world addition
POL	POL_NFZ_REFUND_NC	HEARING_LOSS_H90_H91	claims	sensitivity set	plausible	rest-of-world addition
POL	POL_NFZ_REFUND_NC	F84_AUTISM_PLN	claims	never used	excluded	rest-of-world addition
POL	POL_SIO_OTHER_VOIV	VISION	school_census	primary control	minor	v5 specified list
POL	POL_SIO_OTHER_VOIV	HEARING	school_census	sensitivity set	plausible	v5 specified list
POL	POL_SIO_OTHER_VOIV	MOTOR_OR_APHASIA	school_census	sensitivity set	plausible	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	INTEGRATED_PHYSICAL	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	INTEGRATED_VISION	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_CLASS_PHYSICAL	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_CLASS_VISION	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_SCHOOL_PHYSICAL	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_SCHOOL_VISION	school_census	primary control	minor	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	INTEGRATED_HEARING	school_census	sensitivity set	plausible	v5 specified list

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_CLASS_HEARING	school_census	sensitivity set	plausible	v5 specified list
SVK	SVK_CVTI_OTHER_DISABILITY_KRAJ	SPECIAL_SCHOOL_HEARING	school_census	sensitivity set	plausible	v5 specified list
SVN	SVN_SURS_SEN_OTHER_AGE	PHYSICAL	school_census	primary control	minor	v5 specified list
SVN	SVN_SURS_SEN_OTHER_AGE	VISUAL	school_census	primary control	minor	v5 specified list
SVN	SVN_SURS_SEN_OTHER_AGE	HEARING	school_census	sensitivity set	plausible	v5 specified list
SVN	SVN_SURS_SEN_OTHER_AGE	LONG_TERM_ILL	school_census	sensitivity set	plausible	v5 specified list
SWE	SWE_NPR_CONTROLS	K35	registry	primary control	none	v5 specified list
SWE	SWE_NPR_CONTROLS	Q90	registry	primary control	none	v5 specified list
TWN	TWN_MOE_SPED_NC	TWN_SPED_CP	school_census	primary control	minor	v5 specified list
TWN	TWN_MOE_SPED_NC	TWN_SPED_PHYSICAL	school_census	primary control	minor	v5 specified list
TWN	TWN_MOE_SPED_NC	TWN_SPED_VISUAL	school_census	primary control	minor	v5 specified list
TWN	TWN_MOE_SPED_NC	TWN_SPED_HEALTH	school_census	sensitivity set	plausible	v5 specified list
TWN	TWN_MOE_SPED_NC	TWN_SPED_HEARING	school_census	sensitivity set	plausible	v5 specified list
TWN	TWN_MOHW_DEATHS_COUNTY	TWN_DEATH_ACCIDENTS	deaths	primary control	none	v5 specified list
TWN	TWN_MOHW_DEATHS_TOWNSHIP	TWN_DEATH_ACCIDENTS	deaths	primary control	none	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_ALBINO_INCL_STAGE	school_census	primary control	none	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_PHYSICAL_IMPAIRMENT_INCL_STAGE	school_census	primary control	minor	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_BLIND_INCL_STAGE	school_census	sensitivity set	plausible	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_DEAF_BLIND_INCL_STAGE	school_census	sensitivity set	plausible	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_DEAF_INCL_STAGE	school_census	sensitivity set	plausible	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_HARD_OF_HEARING_INCL_STAGE	school_census	sensitivity set	plausible	v5 specified list
TZA	TZA_BEST_DISABILITY_REGION	TZA_BEST_LOW_VISION_INCL_STAGE	school_census	sensitivity set	plausible	v5 specified list
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_ALBINO	school_census	primary control	none	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_PHYSICAL_IMPAIRMENT	school_census	primary control	minor	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_BLIND	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_DEAF	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_DEAF_BLIND	school_census	sensitivity set	plausible	rest-of-world addition

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_HARD_OF_HEARING	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_LOW_VISION	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_VISUAL_IMPAIRMENT	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_INTELLECTUAL_IMPAIRMENT	school_census	never used	excluded	rest-of-world addition
TZA	TZA_BEST_PREPRIMARY_REGION	TZA_BEST_MULTI_IMPAIRMENT	school_census	never used	excluded	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_ALBINO_INCL_STAGE	school_census	primary control	none	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_PHYSICAL_IMPAIRMENT_INCL_STAGE	school_census	primary control	minor	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_BLIND_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_DEAF_BLIND_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_DEAF_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_HARD_OF_HEARING_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_LOW_VISION_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_VISUAL_IMPAIRMENT_INCL_STAGE	school_census	sensitivity set	plausible	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_INTELLECTUAL_IMPAIRMENT_INCL_STAGE	school_census	never used	excluded	rest-of-world addition
TZA	TZA_BEST_PRIMARY_COUNCIL	TZA_BEST_MULTI_IMPAIRMENT_INCL_STAGE	school_census	never used	excluded	rest-of-world addition
USA	(not written)	developmental delay	school_census	never used	excluded	v5 specified list
USA	(not written)	emotional disturbance	school_census	never used	excluded	v5 specified list
USA	(not written)	intellectual disability	school_census	never used	excluded	v5 specified list
USA	(not written)	multiple disabilities	school_census	never used	excluded	v5 specified list
USA	(not written)	other health impairment	school_census	never used	excluded	v5 specified list
USA	(not written)	specific learning disability	school_census	never used	excluded	v5 specified list
USA	(not written)	speech or language	school_census	never used	excluded	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_OI	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_OI_EC	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_OI_SA	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_TBI	school_census	primary control	none	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_TBI_EC	school_census	primary control	none	v5 specified list
USA	USA_IDEA_PARTB_STATE_NC	IDEA_B_TBI_SA	school_census	primary control	none	v5 specified list

Country	Source	Outcome code	Outcome system	Role	Vaccine pathway	List
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_VI	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_VI_EC	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_VI_SA	school_census	primary control	minor	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_DB	school_census	sensitivity set	plausible	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_DB_EC	school_census	sensitivity set	plausible	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_DB_SA	school_census	sensitivity set	plausible	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_HI	school_census	sensitivity set	plausible	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_HI_EC	school_census	sensitivity set	plausible	v5 specified list
USA	USA_IDEA_PART B_STATE_NC	IDEA_B_HI_SA	school_census	sensitivity set	plausible	v5 specified list

<!-- /S77-A -->

S77.2 The contrast

On matched comparisons (same record system, jurisdiction level, jurisdiction and first affected cohort, same events and uptake paths), the country contrast is the pooled autism step minus the pooled control step, with variance equal to the sum of the two (the autism side keeps the kernel envelope). Countries are pooled with weights 1/(v + 0.33^2). The calibrated interval uses variance components fitted to matched placebo contrasts at offsets seven to twelve: shared 0.00, country 0.200. As the rule requires, the contrast code was run on synthetic counts with no effect, an autism-only per-child effect of 1.25 and a shared per-child bias of 1.15 in both outcomes; the contrast recovered 1.26 and 1.03 respectively, while the autism-only step under the shared bias was 1.18 (code checks, not data).

<!-- S77-B -->

Table S77-B. The contrast and its parts (matched comparisons; random-effects pooling by country).

Quantity	Estimate	95 percent interval
Matched autism step (per-child relative risk)	0.940	
Matched control step	0.998	
Contrast, primary controls, random effects	0.913	0.759-1.098
Contrast, primary controls, placebo-calibrated (sensitivity)	0.913	0.796-1.046
Contrast, sensitivity control set, placebo-calibrated	1.000	0.854-1.171
Unmatched autism-to-control ratio (as in version 4), placebo-calibrated	0.962	0.813-1.138

The contrast pools 49,966 matched comparisons in 25 countries; its effective number of countries is 17.1.

<!-- /S77-B -->

S77.3 New sources

Ecuador's school census (all schools; school-entry age five where grades are mapped), the French RHEOP register (age-7 birth cohorts by département, a family separate from SNDS), Mexican hospital discharges (any-diagnosis F84), Israel's CBS registry, Korean claims at age five and Australian MBS items (paediatrician and psychiatrist as two slots) enter under version 4's rules; they contribute 279 comparisons and 9.0 percent of the headline weight. Modeled U.S. county coverage (NIS small-area estimates) does not qualify a sub-state comparison in the primary, because those estimates are model-based; it does in the specified sensitivity (0.973).

S77.4 Norway and Spain

Norway's coverage paths now include the zeros before each programme's introduction (an exposure-construction fix, not an outcome-driven change). The version 4 Norway estimates had been seen, so the version 4 paths are refitted as a sensitivity: 0.968. Spain's national single-grade table is a disclosed post hoc sensitivity, 0.975; the primary keeps the provinces, the finest level with measured exposure, chosen without outcomes.

S77.5 Review disclosures

The reporting follows two external reviews. Coverage-collapse events lead 9,054 primary comparisons (9.5 percent), 8,765 of them Brazil's 2019 fall in second-dose measles coverage; amendment A3 left them unamended, and dropping them is a post hoc sensitivity (0.974). The other post hoc pool-stage sensitivities the reviews requested give 0.966 with the version 5a countries only, 0.975 without microstates and 0.979 without single-stack countries; an attenuation-corrected point estimate, a disclosure only, is 0.970. The Dutch event labelled MenACWY in 2002 was monovalent MenC; its weight is negligible. The raw-scale autism-to-control ratio, 2.68, is a quotient of two noisy steps near zero and carries no information. The unmatched autism-to-control ratio is 0.962 (placebo-calibrated 0.813-1.138).

S77.6 Data expansion and what was not obtained

A global coverage matrix listed 665 country-by-level rows, in 228 countries, that had not been checked for autism or exposure data at baseline. A second pass re-checked every cell that the first pass had marked as holding no usable source; in the original matrix 94.0 percent of those verdicts survived (1,990 of 2,116 cells), and reopened cells entered the parts queue. Sources that could not be obtained without the author are listed on the D6 docket: 234 de-duplicated items, of which 112 need a data request, freedom-of-information request or application and 122 need a browser download, library access, a free account or a technical unblock. No one was contacted.

S78. Robustness of the per-child conversion (amendment A5)

The author kept the label 'per-child relative risk'. Amendment A5, specified before any of these checks was computed, tests the three things the label rests on: the uptake path p, the mixture model 1 + p(r - 1) and the outcome definition. Every check runs on the sealed headline profiles, at the pool stage or as a stack-level meta-regression, and reports the per-child relative risk, its assumed-allowance interval (random effects plus a shared standard deviation of 0.15), the random-effects interval and the effective number of countries. These are robustness checks, not headline rules. By the rule fixed in A5, the label is supported if the P1-P4 allowance intervals overlap the headline's, the M1 intercept interval includes zero, the M2 tertiles agree and no single-country exclusion (O3) leaves the headline's allowance interval. 6 of the 7 criteria pass; the label is therefore not fully supported, and each failing criterion is a limitation (journal Section 8).

Table S78-A. Uptake-path quality (P1-P4): per-child relative risks by stratum.

Check	Per-child relative risk	assumed-allowance interval	Random-effects interval	Effective K	Countries	Stacks
Headline	0.973	0.699-1.354	0.836-1.131	26.0	40	39,753
P1 measured uptake	0.934	0.629-1.387	0.717-1.217	8.8	14	8,355
P1 measured and modeled (mixed)	0.963	0.670-1.383	0.779-1.190	12.6	16	16,343
P1 modeled uptake	0.967	0.684-1.366	0.807-1.159	20.9	37	15,055
P2a modeled-zero pre-programme cohorts excluded	0.894	0.613-1.303	0.706-1.132	13.2	25	17,247
P2b stacks without a measured pre-change cohort excluded	0.851	0.514-1.409	0.566-1.281	4.6	8	3,864
P3 countries with subnational uptake variation	0.993	0.698-1.412	0.818-1.205	14.9	18	39,532
P3 countries with national uptake only	0.942	0.644-1.379	0.739-1.200	11.4	22	221
P3 broadcast-only countries dropped	0.985	0.706-1.375	0.842-1.153	24.1	38	39,720

Check	Per-child relative risk	assumed-allowance interval	Random-effects interval	Effective K	Countries	Stacks
P4 uptake jump of at least 0.30	0.984	0.706-1.371	0.843-1.148	25.1	38	20,514
P4 uptake jump below 0.30	0.921	0.619-1.370	0.705-1.203	9.5	22	19,177

The measured and modeled strata do not differ (Q = 0.04 on 1 df, p = 0.83). P5, the measurement-error band: if the true uptake were 0.70 to 1.30 times the uptake used, the per-child relative risk would range from 0.961 to 0.979, with allowance limits from 0.599 to 1.541.

Table S78-B. The mixture model (M1-M3).

Check	Estimate	95 percent interval
M1 intercept, random country intercept	+0.180	+0.096 to +0.264
M1 intercept, country-cluster robust	+0.180	+0.068 to +0.293
M1 intercept, random country slope added (post hoc)	+0.095	-0.015 to +0.206
M1 intercept, Brazil excluded (post hoc)	+0.126	+0.040 to +0.212
M1 intercept, coverage-collapse leaders excluded (post hoc)	+0.147	+0.062 to +0.232
M1 intercept, country and lead-event effects (post hoc)	+0.099	+0.023 to +0.174
M1 implied r, random country intercept	0.797	0.787-0.807
M1 implied r, country and lead-event effects (post hoc)	0.959	0.924-0.994
M1 implied r, random country slope added (post hoc)	0.952	0.785-1.119
M2 lowest tertile of the uptake change	0.895	0.563-1.421
M2 middle tertile	0.910	0.620-1.336
M2 highest tertile	0.982	0.704-1.370
M3 exact mixture inversion, headline	0.973	0.687-1.342
M3 log-linear approximation, headline	0.973	0.699-1.354

M1 regresses stack-level raw log-steps on the uptake change with a country random effect (the specified model); under the mixture model the intercept is zero and the slope gives r (implied r = 1 + slope). Rows marked post hoc were added after the specified model was fitted; they are diagnostics and do not change the verdict. The model with an intercept fits better (likelihood ratio 14.3, AIC lower by 12.3). M2 intervals are assumed-allowance intervals; the tertiles agree (Q = 0.40 on 2 df, p = 0.819). M3 intervals are assumed-allowance intervals.

Table S78-C. Outcome system, product of the lead event and single-country exclusions (O1-O3).

Check	Per-child relative risk	assumed-allowance interval	Random-effects interval	Effective K	Countries	Stacks
O1 school census	0.941	0.660-1.340	0.772-1.146	16.3	25	5,840
O1 claims or outpatient	1.106	0.697-1.757	0.774-1.582	4.4	7	22,308
O1 hospital	0.819	0.456-1.470	0.494-1.358	3.0	4	472
O1 registry	0.989	0.658-1.487	0.746-1.312	8.0	13	410
O1 benefits	1.094	0.641-1.868	0.700-1.710	2.6	3	10,719
O2 PCV	0.926	0.648-1.324	0.756-1.135	15.2	28	4,385
O2 MMR or MCV	0.770	0.411-1.442	0.442-1.341	2.8	9	6,425
O2 rotavirus	1.055	0.716-1.557	0.819-1.361	10.8	21	2,266
O2 hepatitis B	1.064	0.693-1.632	0.779-1.452	7.4	23	755
O2 Hib or pentavalent	1.027	0.685-1.540	0.778-1.357	9.7	19	452
O2 varicella	0.990	0.616-1.590	0.682-1.436	4.7	8	4,979

Check	Per-child relative risk	assumed-allowance interval	Random-effects interval	Effective K	Countries	Stacks
O2 other	0.920	0.633-1.336	0.731-1.157	12.3	23	20,439
O3 Brazil dropped	0.971	0.697-1.355	0.832-1.135	25.2	39	6,888
O3 United States dropped	0.977	0.701-1.362	0.836-1.141	25.2	39	35,673
O3 Britain dropped	0.973	0.698-1.357	0.833-1.136	25.2	39	39,131
O3 Peru dropped	0.970	0.696-1.353	0.831-1.134	25.2	39	39,732
O3 Korea dropped	0.977	0.700-1.362	0.836-1.141	25.2	39	39,592

Heterogeneity by outcome system: Q = 1.33 on 4 df, p = 0.86; by product: Q = 2.01 on 6 df, p = 0.92.

Table S78-D. Planted-effect recovery by stratum (G): recovered per-child relative risk and log-scale error for planted 1.25 and 1.5.

Stratum	Recovered (1.25)	Error, percent	Recovered (1.5)	Error, percent
Headline	1.227	-8.3	1.460	-6.7
P1 measured uptake	1.187	-23.2	1.378	-20.9
P1 mixed	1.221	-10.4	1.443	-9.5
P1 modeled uptake	1.241	-3.1	1.491	-1.4
O1 school census	1.230	-7.1	1.466	-5.7
O1 claims or outpatient	1.209	-14.9	1.419	-13.7
O1 hospital	1.218	-11.7	1.453	-7.9
O1 registry	1.242	-2.8	1.495	-0.9
O1 benefits	1.205	-16.3	1.418	-13.8

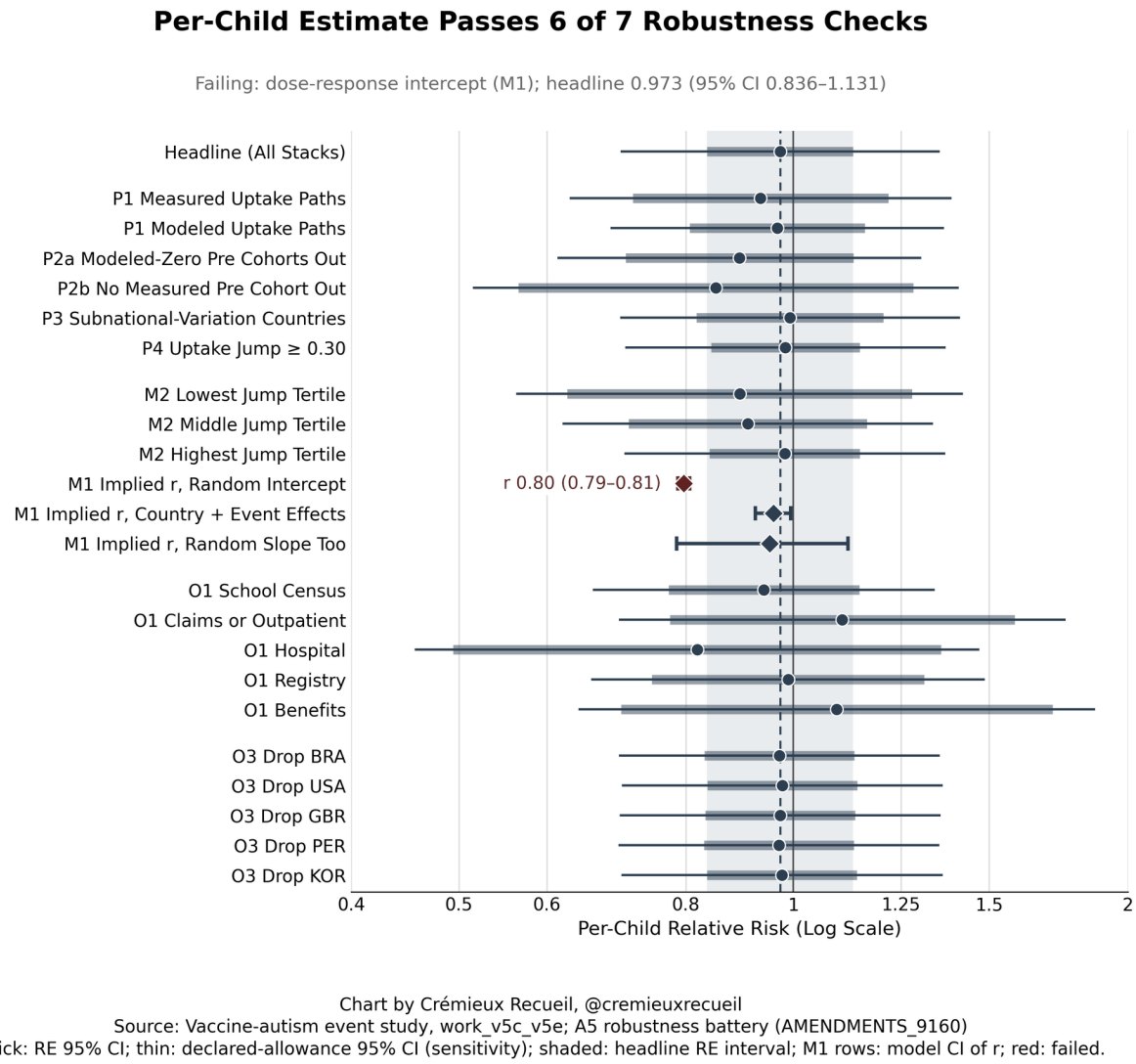


Figure S123. The A5 robustness battery for the per-child conversion: per-child relative risks with 95 percent assumed-allowance intervals by uptake-path stratum (P1-P4), tertile of the uptake change (M2), outcome system (O1) and single-country exclusion (O3), with the r implied by the dose-response regression (M1). Robustness checks specified before they were computed, not headline rules.

S79. Results with all held data (maximal inclusion, amendment A6)

Amendment A6, specified before it was computed, refits the headline design with every held series the design can place on birth cohorts: short series, survey coverage as paths, the measured-level promotion and outcome parts that are not duplicates of a primary part (with a source fixed effect). Duplicates of a primary source stay out, and parts without any age, grade or cohort dimension cannot be placed. The fit is report-only: it has no headline role, whatever its diagnostics show.

Table S79-A. The headline and the maximal-inclusion fit.

Fit	Per-child relative risk	Random-effects 95 percent CI	Placebo-calibrated (sensitivity)	assumed allowance (sensitivity)	Effective K	Countries	Stacks
Headline	0.973	0.836-1.131	0.855-1.107	0.699-1.354	26.0	40	39,753
Maximal inclusion	0.961	0.835-1.107	0.856-1.079	0.694-1.332	28.8	44	48,939

EA1. Recovery under the native perturbation

The main result and its recovery experiment answer different questions. Figure EA1 shows the observed event-time estimates and the native pipeline after deterministic perturbations along the centered uptake path. The target labels 1.25 and 1.50 name the injected perturbations. They are not independently observed receipt effects.

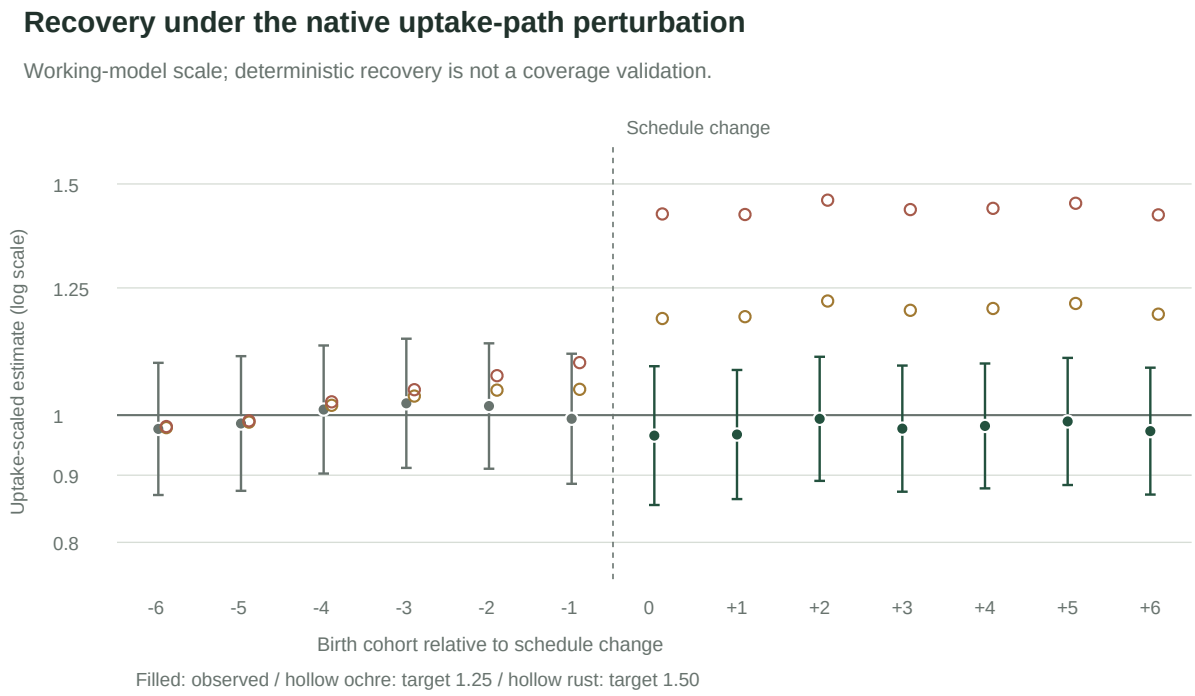


Figure EA1. Observed event-time estimates with native injected-path scenarios. Filled points show observed data with 95 percent random-effects intervals. Hollow points show the native pipeline after each deterministic perturbation. The source-stratum recovery failures are retained in Section S78. All values use the original public A7 analysis.

The pooled recovered ratios are 1.23 and 1.46, with log-target errors of -8.3 and -6.7 percent. Recovery is attenuated under these scenarios. An attenuation-adjusted point of 0.970 was reported as a disclosure in the original paper; it is not adopted as an alternative causal estimate. The experiment does not validate absolute baseline receipt, counterfactual comparability or repeated-sample interval coverage. Four of six ramp placebo intervals miss zero, and measured-uptake recovery fails the specified 10 percent standard (Section S78).

EA2. Analysis editions and publication provenance

The main paper and Figure 13 use the original public v5e/A7 analysis snapshot of 15 October 2026. The manuscript date is 28 October 2026. This editorial revision consolidates figures and moves development history to this appendix. It does not run new source fits or substitute the later private timing-qualified candidate.

Analysis edition	Estimate	Random-effects interval
Fourth-edition reference	0.965	0.810-1.150
Expanded v5a reference	0.963	0.820-1.130
v5c, amendment A2	0.972	0.833-1.134
v5c, A3 corrections	0.973	0.834-1.136
Public v5e/A7 snapshot	0.973	0.836-1.131

These rows reuse populations, outcome series and analysis code. They document development and are not independent replications. The v5a matched contrast failed its injection gate. The Norway exposure-path sensitivity gives 0.968 and the additional Spanish single-grade sensitivity gives 0.975; the latter was specified after the earlier result had been seen. Section S77 documents the source corrections.

The v5 rule (RULE 9160) was pinned in execution freeze 4977 before the first model call on 10 October 2026 at 22:03 CDT. Earlier results had already been inspected, and companion hash files promised by the rule were not written. The A2 amendment was written before its fits, while the short-calendar-span restriction was motivated by preview diagnostics. These records document timing of choices; they do not constitute external preregistration. The model-selection and diagnostic limitations remain in the main paper.

The original journal Figure 13 was re-rendered through a release script that lacked a v5e/A7 label branch and printed v5c/A2 in its footer. The consolidated figure is generated directly from the retained A7 chart and matched-summary values. The shared-bias table (main Table 5) uses the same verified A7 estimate and variances as the text, replacing the stale fourth-edition table. Table 3 labels the kindergarten-MMR association as fourteen states, matching the saved fit membership. The point estimate and interval for that association are unchanged.

Main Table 5 uses shared SD scenarios 0.00, 0.08 and 0.15; the first two correspond to the step-placebo estimate and its upper profile endpoint, and the third is the assumed allowance. Its intervals are arithmetic sensitivity calculations from retained estimates and variances. They are not a new empirical coverage validation. The original table used the historical 27-country result and the older 0.09 endpoint; that historical analysis remains in S74.

EA3. Full matched-control event-time display

Figure 13B gives the compact pooled matched comparison. Figure EA3 preserves its birth-cohort display. It uses 49,966 matched comparisons in 25 countries; event-time membership differs by cohort. The matched result is not the simple average of the displayed cohort points.

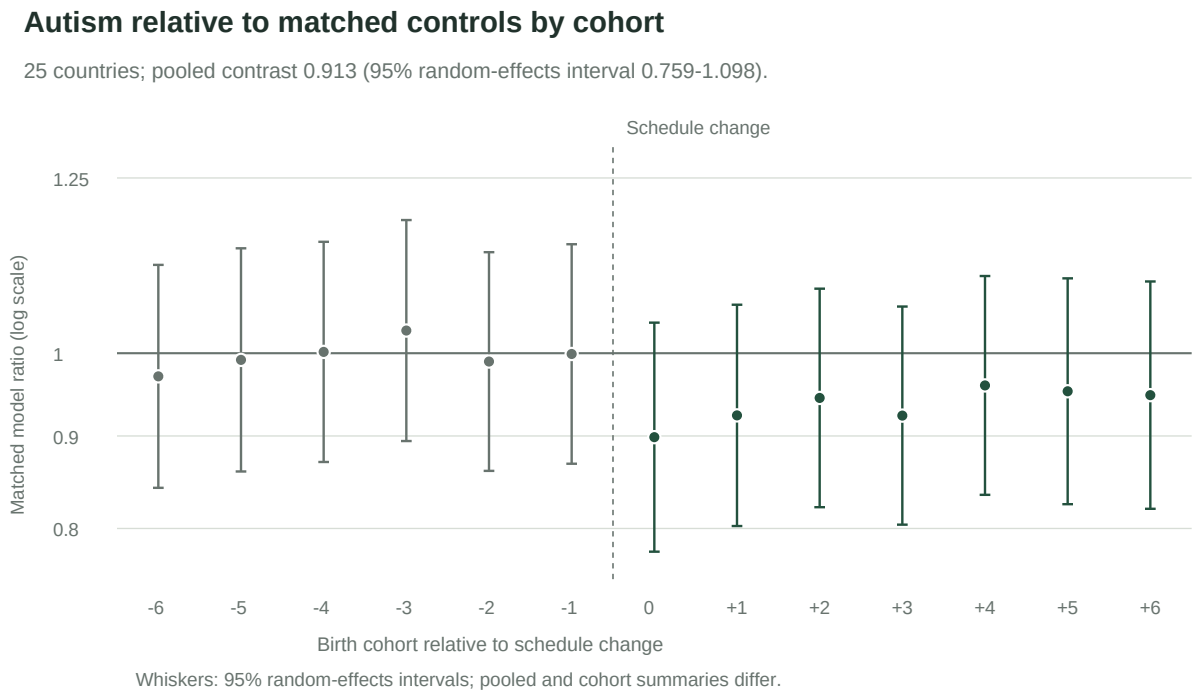


Figure EA3. Matched autism-minus-control estimates by birth cohort relative to each schedule change. Whiskers are 95 percent random-effects intervals. The pooled estimate is 0.913 (0.759-1.098); the placebo-calibrated sensitivity is 0.796-1.046. The matched observations overlap the primary event study and are not independent confirmation.

Matching can remove only the outcome-side biases that autism shares with the selected controls in the same record system. An autism-specific bias may remain, and receipt measurement is not validated by this comparison. The sensitivity control set gives 1.000, the unmatched ratio 0.962, autism alone on matched observations 0.940 and controls alone 0.998. These are distinct contrasts on overlapping data.

EA4. Supporting results and numerical verification

This section supplies the detailed tables supporting the main findings. The scientific estimates use the original public A7 analysis and the complementary models identified in each table. Counts refer to overlapping analyses, not independent experiments. Numerical reproduction, source provenance and causal identification are separate questions. No scientific assumption or diagnostic threshold has been changed to improve agreement.

Table EA4.1. Evidence inventory

Evidence layer	Scope	Interpretation
Combined result register	77,738 records	Coefficients, diagnostics and repeated specifications; not independent studies.
Original-observation autism model	17 contexts; 21 contrasts; 16 blocks	8,538 cells; conditional heterogeneous shared mean, not universal individual risk.
International original-panel likelihoods	101 blocks retained	Autism and other outcomes kept distinct; failed and weak diagnostics remain visible.
Local coverage gradients	22 designs; 21 states; 7 families	Special-education composition separated from enrollment-based rates; MMR distinguished from schedule compliance.
Published-source results	4,713 contrasts; 87 participant/source identifiers	37 overlapping native-measure families and sensitivity fits; identifiers include shared and composite sources.
Interactions and cumulative load	27 interaction contrasts; 3 working overlap groups	Interaction multiplier and schedule-load slope answer different questions; load remains weakly informed.

Table EA4.2. Full native-scale estimates and identification diagnostics

Setting	Effect unit	RR	95% interval	p	Pretrend p	Identification
Japan: MMR coverage	RR per +10pp coverage	0.981	0.805-1.196	0.85	n/a	limited: nine aggregate cohorts
Spain: varicella convergence	late-vs-early post RR	0.573	0.296-1.112	0.0942	0.0197	failed: differential pretrend
U.S. RotaShield pulse	RR per +10pp uptake	0.996	0.945-1.051	0.895	n/a	limited: only one clean lead
U.S. PCV7 rollout	RR per 1-SD early uptake	1.013	0.981-1.046	0.43	0.636	Ranking contrast only; nonmonotone exposure mapping; no individual-risk exclusion (S52)
U.S. PCV7 diagnostic-substitution DDD	autism-specific differential RR	0.995	0.952-1.040	0.824	0.933	supportive falsification design
U.S. SchoolVax MMR dose-response	RR per +10pp kindergarten MMR coverage	0.989	0.930-1.052	0.718	n/a	limited: only three pandemic-adjacent cohorts
U.S. ADDM PCV7 replication	RR per 1 six-site SD (2.35 pp) early uptake	1.053	0.971-1.143	0.262	0.149	independent but small-cluster clinical-outcome replication
England: rotavirus, pooled ages	eligible-vs-pre RR	0.940	0.891-0.993	0.0282	n/a	fragile: APC/trend sensitivity
Scotland: rotavirus ramp	full-share RR per 1-SD uptake	1.020	0.567-1.833	0.946	0.137	limited: proxy intensity
Wales: rotavirus ramp	full-share RR per 1-SD uptake	0.875	0.489-1.566	0.639	0.0102	failed: pretrend/system break
Northern Ireland: rotavirus	fully exposed-vs-pre RR	1.078	0.915-1.271	0.37	1.51e-97	failed: APC/pretrend
Sweden: PCV rollout	fully exposed-vs-pre RR	0.886	0.708-1.109	0.292	1.39e-07	failed: pretrend/outcome mismatch
U.S. kindergarten MMR; 14 states	Autism share of special education / +10 pp	1.002	0.950-1.057	not used	mixed diagnostics	95% credible interval with assumed bias allowances; composition association, not population autism risk
Brazil municipality cohorts	RR / +10 pp	1.000	0.992-1.008	0.991	younger/older windows null	Exploratory: one Census, five-year windows, ecological PNI coverage
Peru CONADIS autism	RR / +10 pp	1.759	1.093-2.829	0.0201	future exposure fails	Excluded: service composition, registration timing, nonspecific outcomes
Overlap-safe PCV7 introduction-kink synthesis	RR / cohort-slope kink	0.983	0.928-1.043	0.429	n/a (kink)	specified age; mutually exclusive CA/MN/NJ/Japan panels; Japan 53.9% of weight (SE calibrated x3.14); without Japan 0.962 (0.857-1.081); federal IDEA excluded for overlap

Setting	Effect unit	RR	95% interval	p	Pretrend p	Identification
U.S. PCV7 state-specific kink replication	RR / cohort-slope kink	0.991	0.973-1.009	0.315	|z| audit 0.80	All 51 jurisdictions; same EDFacts source at finer grain, not independent evidence

Rate ratios use row-specific effect units and are not pooled. A favorable p-value does not override a failed pretrend, placebo, measurement, or age-period-cohort diagnostic. Local MMR: denominator-specific Bayesian ratio, with a 95% credible interval.

Table EA4.3. Deterministic effect recovery by source stratum

Stratum	1.25 recovered	Log error	Gate	1.50 recovered	Log error	Gate
Primary model	1.227	-8.3%	Pass	1.460	-6.7%	Pass
Measured paths	1.187	-23.2%	Fail	1.378	-20.9%	Fail
Modeled paths	1.241	-3.1%	Pass	1.491	-1.4%	Pass
Mixed paths	1.221	-10.4%	Fail	1.443	-9.5%	Pass
School census	1.230	-7.1%	Pass	1.466	-5.7%	Pass
Claims or outpatient	1.209	-14.9%	Fail	1.419	-13.7%	Fail
Hospital records	1.218	-11.7%	Fail	1.453	-7.9%	Pass
Registry	1.242	-2.8%	Pass	1.495	-0.9%	Pass
Benefits and service records	1.205	-16.3%	Fail	1.418	-13.8%	Fail

The gate requires absolute error within 10 percent of the planted log target. Ten of eighteen stratum-by-target checks pass and eight fail. The primary perturbations return 1.227 and 1.460 for targets of 1.25 and 1.50. This tests recovery of specified deterministic perturbations; it does not establish receipt accuracy, causal validity or interval coverage. The original receipt-conversion battery passes six of seven criteria; its zero-intercept dose-response criterion fails (+0.180; 95 percent interval +0.096 to +0.264).

Table EA4.4. Placebo and holdout diagnostics

Check	Observed result	Interpretation
Pooled fake-date intervals	6 / 6 include zero	Pooled check on reused source data
Country fake dates; sampling only	150 / 187 (80.2%)	Below the nominal 95% rate
Fitted country-bias calibration	180 / 187 (96.3%)	In-sample calibration
Held-offset calibration	180 / 187 (96.3%)	Six held-date splits; overlapping source records
Held-country calibration	180 / 187 (96.3%)	34 held-country splits; Czechia and Taiwan retain misses
Ramp fake-date intervals	2 / 6 include zero	Four exclude zero; gradual +25% remains compatible

Held-offset and held-country calculations reuse the same source records across splits. The observed rate is not an independent certificate of future 95 percent coverage. Country-level misses concentrate in Czechia and Taiwan; joined metadata and cached summaries agree, but this does not establish raw-source fit or covariance parity.

Table EA4.5. Exhaustive country deletions

Deletion	Analyses	Point-estimate range	Interval result
Omit any 1 country	40	0.965 to 0.985	All random-effects intervals include 1
Omit any 2 countries	780	0.957 to 0.993	All random-effects intervals include 1

Table EA4.6. Stipulated cross-country dependence

Assumed correlation scenario	Estimate	95% interval
Correlation 0.10 on within-country variance	0.973	0.787 to 1.202
Correlation 0.10 on all marginal variance	0.973	0.716 to 1.321

These are two examples from 32 stipulated covariance scenarios, not estimates of cross-country dependence. No extra shared-bias component is included in these two rows. The assumed shared log-bias SD of 0.15 already makes a 25 percent increase compatible with the primary estimate.

Table EA4.7. Outcome system

Outcome system	Ratio	95% RE	95% shared-bias	Countries	Comparisons
School census	0.941	0.772 to 1.146	0.660 to 1.340	25	5,840
Claims or outpatient	1.106	0.774 to 1.582	0.697 to 1.757	7	22,308
Hospital records	0.819	0.494 to 1.358	0.456 to 1.470	4	472
Registry	0.989	0.746 to 1.312	0.658 to 1.487	13	410
Benefits and service records	1.094	0.700 to 1.710	0.641 to 1.868	3	10,719

Table EA4.8. Leading schedule-event product

Leading schedule-event product	Ratio	95% RE	95% shared-bias	Countries	Comparisons
Pneumococcal	0.926	0.756 to 1.135	0.648 to 1.324	28	4,385
MMR or measles-containing	0.770	0.442 to 1.341	0.411 to 1.442	9	6,425
Rotavirus	1.055	0.819 to 1.361	0.716 to 1.557	21	2,266
Hepatitis B	1.064	0.779 to 1.452	0.693 to 1.632	23	755
Hib or pentavalent	1.027	0.778 to 1.357	0.685 to 1.540	19	452
Varicella	0.990	0.682 to 1.436	0.616 to 1.590	8	4,979
Other lead-event products	0.920	0.731 to 1.157	0.633 to 1.336	23	20,439

Strata use their own filters and share countries and records. Lead-event product strata are not isolated clinical product effects. Shared-bias intervals use the assumed allowance.

Table EA4.9. Common-effect and schedule-load models

Model	Estimate	95% credible interval	Basis
Common effect; assumed bias	0.964	0.637 to 1.455	15 contexts; 5,133 cells
Common effect; doubled bias	0.977	0.540 to 1.763	Same model with larger bias allowances
Control-calibrated common effect	0.871	0.741 to 1.020	Requires control-to-autism bias exchangeability
Control calibration; residual SD 0.15	0.876	0.632 to 1.212	Mean-shift calibration plus specified residual bias
Antigen load; per five antigens	0.999	0.507 to 1.965	Prior-dominated; ten schedule-load contexts
Injection load; per five injections	0.982	0.516 to 1.865	Prior-dominated; same contexts

Table EA4.10. Selected literature-family models

Family	Measure	Contrasts	Estimate	95% credible interval
Childhood MMR and autism	HR	2	0.954	0.647 to 1.415
Prenatal influenza and autism	HR	4	1.012	0.779 to 1.315
MMR and autism; case-control	OR	2	0.926	0.585 to 1.513
PCV13 and febrile seizures; days 0-1	IRR	3	1.770	1.116 to 2.720
Influenza plus PCV13; febrile seizures	IRR	2	3.027	1.413 to 5.403
Rotavirus and type 1 diabetes; VSD	HR	2	1.093	0.652 to 1.805
Rotavirus and type 1 diabetes; MarketScan	HR	2	1.107	0.699 to 1.738
Rotavirus and type 1 diabetes; Optum	HR	2	0.874	0.534 to 1.533
Rotavirus and coeliac disease	HR	3	1.032	0.750 to 1.417

These are bias-adjusted project family summaries, not the published studies' unmodified headline estimates. Families reuse cohorts. Positive febrile-seizure associations are retained. The literature and local reference layers contain 44 reference families and 446 alternative-family rows; archived local membership is distinct from the 14-state MMR result in the main paper.

Table EA4.11. Other outcome findings

Outcome	Estimate	95% interval	Scale and interpretation
All malignant neoplasms	1.00	0.78 to 1.28	13 gated contexts; model-implied ratio
Leukaemia	1.07	0.79 to 1.43	Project model; moderate effects remain compatible
Lymphoid leukaemia	1.16	0.83 to 1.62	Project model; low precision
Norway type 1 diabetes trend	1.11	1.03 to 1.20	Published relative HR for a slope change; cutoff concerns
Danish ASD high versus middle aluminum	+8.68	+4.98 to +12.38	Published risk difference per 10,000 at age 5; observational

The Norway result is a birth-cohort slope contrast in relative hazards, not a level risk ratio. Excluding 2013-2015 yields 1.05 (0.93-1.18), and other policy cutoffs show similar breaks. The Danish result is an observational risk difference at age five for high versus middle aluminum exposure, not the Cox-model exposure slope; the full set of contrasts does not show a coherent monotone relation. Cancer and chronic-outcome rows retain their outcome-specific limits and receive no causal-synthesis weight under the project criteria.

EA4.12. Numerical reproduction and source scope

Verification scope

Component	Extent	Status
Primary country pool	40 countries; 104 specification pools	Recomputed from saved country fits
Country deletion robustness	40 single and 780 pair deletions	All retained random-effects intervals include 1
Reference-family synthesis	37 literature and 7 local families	44 native reference models refitted
Older joint hierarchy	120 probability scenarios	Archived likelihood grid reintegrated
Swedish service-use models	120 main and 40 event-time fits	Independent coefficient and cluster-SE checks agree
Assignment and native power	4,200 contrasts; 120 power values	Independent arithmetic agrees
Swedish pair covariance	912 covariance recalculations	Current arithmetic independently validated
Historical placebo SE equality	Four displays; two distinct inputs	Unresolved; original reference values retained
Source-document extraction	Only supplied panels replayed in this pass	Full raw-source replay not certified
Later timing-qualified candidate	Separate reported result	Not substituted into this public-analysis summary

The public release manifest contains 1,844 files whose hashes were verified. Saved country-level inputs reproduce all 104 specification pools; 44 reference-family models and 120 older-hierarchy probability scenarios were replayed. These checks do not independently reacquire every source or validate every model assumption.

For the Swedish service-use analyses, 120 main fits, 40 event-time fits, 4,200 assignment contrasts and 120 native conditional-power values were reproduced. Independent fixed-effect residualization and cluster-standard-error calculations agree across all 160 fits. The checks cover 5,944 saved numeric fields. Scientific diagnostics remain unfavorable: 32 of 40 overlapping pre-period tests have nominal p below 0.05, including all 20 log-outcome tests, and 28 of 40 panel-window-scale groups change coefficient sign across original specifications. Only two counties are treated. These service-use outcomes are neither autism incidence nor individual risk.

Source reconciliation retains 3,722 rows matched on both keys, including nine true zero cells. Nonresponse cells are excluded. Of the matched decimal rates, 2,875 differ in some digit (including precision-only differences), 334 differ by more than one per 10,000 and 114 of the latter occur in model years 2012-2019. Context receipts preserve large discrepancies without attributing them to a source revision. Three of 42 missing upstream paths have hash-identical local counterparts; the other 39 original paths remain absent. Full raw-source extraction is not certified.

For Swedish pair-model covariance, 3,984 fit calls and 912 covariance recalculations across three batteries retain bitwise-identical fitted states. Stable QR covariance uses the original estimator, data and ridge of 1e-8 times max(1, mean fitted count). Independent 90-digit calculations validate current standard errors of 217.040318416916 and 25.025362405184 for the two distinct problematic placebo states. The earlier printed values are 217.040561920589 and 25.025361864064. Four displays refer to these two input states; equality to the historical numbers remains UNVERIFIED because the original fit state and pinned historical runtime are unavailable. Current numerical agreement does not resolve the historical mismatch.

Thirty-two specified row permutations give QR standard errors within 8.3e-12 of the reference. Alternative scoring experiments are diagnostic only and have not replaced the fitted estimator. Flat-information leverage is reported as unavailable for 4, 4 and 5 rows in the three respective variants; the finite-model gates are unchanged. The self-contained current-fit capsule validates the current arithmetic only. The stock scientific diagnostics are not classified as passing.

Additional public-source checks retain their original scope: Pennsylvania and New Jersey count/rate checks agree with the supplied panels; the UAE extension and Colorado/Rhode Island file checks reproduce the archived hashes under explicit original text and listing conventions. A narrow damaged-footer parser repair preserves the UAE numerical

content. Cached Czechia/Taiwan joins and Taipei age-band geometry explain saved metadata or profile behavior, without claiming a fresh primary-source fit.

The displayed estimates retain the original public A7 sample. The retained timing audit reports 42 contributing California configurations with outcomes preceding the recorded date of the selected policy; their exact membership correction remains unverified locally. The later candidate is not substituted. EA6 explains this eligibility limitation and the separate secondary-event holds. Current numerical agreement, source replay and historical equality remain limited to the supplied evidence.

EA4.13. Additional uncertainty calculations

The aggregate event study has its own limits. Its calibrated interval treats bias common to all countries as nearly zero, because the placebo dates estimate the shared component at 0.00 (upper end 0.08). The calibrated country component is 0.260. Under the assumed random-effects model, the country component of 0.33 averages down across about 26 effective countries (0.33/5.10, about 0.065); a bias shared by every country does not average down. Excluding a 25 percent increase therefore relies on the placebo evidence that no large shared bias exists; at the assumed allowance of 0.15 the upper limit is 1.354 (main Table 5). An approximate power benchmark gives 70 percent power at a per-child relative risk of 1.25; 80 percent power requires 1.28 (1.63 with the shared allowance). The diagnostic ledger records 9,054 comparisons led by coverage-collapse events, 8,765 of them Brazil's 2019 fall in second-dose measles coverage; omitting these events is a post hoc sensitivity that gives 0.974 (allowance interval 0.700-1.356). The matched contrast cancels only bias that autism shares with its controls in the same record system, relies on the specified list excluding categories that absorb autistic children, and uses state rather than sub-state series in the United States. Raw steps are not proportional to the size of the uptake jump: regressing them on the uptake change leaves a positive intercept of +0.180 (95 percent CI +0.096 to +0.264; slope -0.203, implied r 0.797), so the per-child scaling rests partly on the proportionality assumption. Post hoc models show that the dose-response is model-dependent: with country and lead-event effects the implied r is 0.959 (0.924-0.994), and with random country slopes the intercept interval includes zero (-0.015 to +0.206).

EA5. Additional model comparisons and diagnostics

These analyses use the models and assumptions identified below. They overlap the main analyses and supply sensitivity checks rather than independent replications.

EA5.1. Additional context models

Smaller benchmarks answer different questions. Intervals in this paragraph are 95 percent credible intervals. Joint PCV/RotaShield fits on 353 U.S. cells plus nine Japanese MMR cohorts yield an equal-baseline-risk standardized average RR 1.012 (0.677-1.631); its 240 assumption combinations concern only those three contexts. A twelve-context, eleven-block model uses 3,338 cells, replacing overlapping Quebec surveillance with age-specific observations, and gives a conditional mean log-effect ratio about 1.02 (0.61-1.66). Replacing Scotland's imputed counts with reported counts gives 1.03 (0.61-1.67) across 3,316 cells. These broad intervals constrain neither every product nor a universal schedule risk; the hypermean differs from the benchmark's arithmetic average.

EA5.2. Control calibration and residual-bias sensitivities

The control-calibrated synthesis estimates bias from 45 candidate controls in nine pooled contexts, assuming their biases are exchangeable with autism's (Supplement S62.9). The controls' mean bias is +0.042 with posterior SD 0.091. The calibrated common effect is 0.871 (0.741-1.020), and the shared mean is 0.877 (0.696-1.102). Common effects of 1.25 or more are excluded in every stated calibration variant, including variance-only calibration, 0.946 (0.806-1.111), and the post hoc removal of Colombia's pandemic school years, 0.943 (0.802-1.105). Exclusions of 1.05 and 1.10 are specification-dependent. The five injury controls of the preview set instead give 1.064 (0.711-1.564). The calibration interprets the U.S. PCV controls' lean as drift shared with autism, moving that context from 1.17 (0.97-1.39) to 0.89 (0.63-1.24), a reading its MIXED diagnostic did not establish. Its minimum local importance-sampling effective size is 16; agreement between two scrambles is not a numerical error bound. P(RR > 1.05) is 0.010 with mean shifts, 0.102 without them and 0.343 under the assumed allowances (S62.12). Controls cannot measure an autism-specific shared error. Allowing additional log-bias SD 0.15 gives 0.876 (0.632-1.212) with mean shifts or 0.949 (0.685-1.314) without them; P(RR > 1.25) is 0.016 or 0.049. At SD 0.30 the upper limits are 1.551 and 1.665 (post hoc; S65; file 3392). These residual-bias values are assumptions, not estimates. Exclusion of a common 25 percent increase is conditional, and medians below one do not establish protection.

EA5.3. Country weighting and the unscaled population step

Comparisons are combined within countries before pooling across countries. Twenty-eight countries carry more than 1 percent of the weight, and the effective number is 26.0. Brazil supplies 83 percent of comparisons and 5.4 percent of the random-effects weight. Alternative information weights give 0.967, with a placebo-calibrated interval of 0.739-1.265 and an outer interval of 0.473-1.974; Brazil, the United States and Britain then receive 45.4, 19.4 and 17.7 percent of the weight. The sampling-only interval, 0.954-0.979, understates the spread seen at placebo dates. The raw population step is +1.7 percent (-7.0 to +11.3); it can differ in sign from the uptake-scaled result because each comparison has its own uptake change and weight.

EA5.4. Effect-recovery checks

For the recovery check, the native pipeline's specified uptake-path perturbations with target ratios of 1.25 and 1.5 return 1.23 and 1.46: log-target errors of -8.3 and -6.7 percent. This demonstrates recovery of that deterministic perturbation, with attenuation under the test scenario; it does not validate absolute receipt, causal interpretation or interval coverage. Recovery is poorer in some source strata (Supplement S78 and EA1). All six pooled fake-date intervals cover zero (largest absolute z 1.20), whereas sampling-only country intervals cover 80 percent of 187 placebo estimates.

EA5.5. Matched-contrast component and control-set sensitivities

The matched autism-versus-control contrast is 0.913 (random-effects interval 0.759-1.098; placebo-calibrated 0.796-1.046; Figure 13B). Its own recovery errors are -8.3 and -6.8 percent and its pooled placebo intervals cover zero at six of six dates. The sensitivity control set gives 1.000 and the unmatched ratio 0.962. On the matched comparisons, autism alone gives 0.940 and controls alone 0.998. These are overlapping analyses, and the contrast removes only biases shared by autism and its controls; an autism-specific bias can remain.

EA5.6. Source and eligibility sensitivities

The estimate is 0.972 and 0.973 at the two endpoints of the between-country standard deviation's interval. Omitting Brazil's outpatient records gives 0.974 (0.837-1.133), omitting mandates gives 0.970, and alternative U.S. exposure and Brazilian case codings give 0.974-0.975 (Supplement Figure S117). School-entry-only exemption laws do not validate the uptake-scaled design: their mean uptake change is 0.032, their recovery check fails, and their raw step is imprecise (+2.0 percent, -22.9 to +34.9). Additional source-path and historical sensitivities are retained in Supplement S74 and EA2.

EA5.7. Receipt-conversion diagnostics and maximal inclusion

The receipt-conversion battery passes six of seven specified criteria (Supplement S78). Measured-uptake, subnational, non-zero-filled and large-jump sensitivity intervals overlap the headline allowance interval; large jumps alone give 0.984 (0.706-1.371). The raw-step regression has a positive intercept of 0.180 (0.096-0.264), contradicting its zero-intercept check. Recovery errors are -23.2 percent on measured-uptake comparisons and -3.1 percent on modeled-uptake comparisons. The individual-risk interpretation is therefore conditional. The report-only maximal-inclusion analysis gives 0.961 (0.835-1.107; Supplement S79).

EA5.8. Uptake-path assumptions and source-specific limits

An abrupt uptake change predicts a cohort step under the stable-effect, comparable-age model. The separate cohort-slope sensitivity is not a within-person accumulation mechanism. School-entry exemption laws fail their recovery check and do not validate the exposure conversion. The per-child scale depends on the uptake paths: in 62 of 134 audited U.S. comparisons, pre-change uptake is a modeled endpoint (no receipt before a recommendation), not measured nonvaccination, so catch-up vaccination would make those conversions conditional. Brazil's outpatient records are included provisionally; omitting them gives 0.974. The PCV-specific sensitivity, 1.00 (placebo-calibrated 0.71-1.42), uses fixed-information pooling (Supplement S74.10). Registry and claims extracts that require accreditation, such as German statutory health-insurance claims, Japanese national claims and Nordic registry tables, would add the most information (Supplement S74); none was used.

EA6. Interpretation of model estimates and diagnostic checks

This guide defines the interpretation of results throughout the paper and appendix, including the retained reference analyses. Numerical estimates, inclusion rules, priors, covariance assumptions and diagnostic thresholds retain their stated definitions. A fitted model parameter, a numerical verification and a causal conclusion answer different questions.

EA6.1. Cohort steps, follow-up age and exposure history

The event-study axis is birth cohort. Under a stable effect and comparable follow-up ages, an abrupt uptake change predicts a cohort step relative to the counterfactual trend. An effect that develops within a person does not by itself produce a ramp across successive birth cohorts. Gradual uptake or a changing effect requires its own specified model. A downward departure after withdrawal can coexist with rising absolute counts if background identification rises. Calendar-year totals mixing cohorts require a separate observation-time mapping.

EA6.2. Exposure scaling and measurement error

The homogeneous mixture is $q0[1 + p(r - 1)]$. For high and low uptake, the exact model ratio is $[1 + pH(r - 1)]/[1 + pL(r - 1)]$. The native event study retains its stated log-linear approximation; these definitions do not refit or correct its results. Classical exposure error can attenuate a coefficient, but systematic, correlated or nonlinear error can bias it in either direction. Recovering about 92 percent of a chosen deterministic perturbation does not establish a general attenuation factor for the real effect or its interval. The failed zero-intercept dose-response check remains a failed check.

EA6.3. Null-compatible intervals, probabilities and model comparisons

An interval containing the null is compatible with no association under the fitted model; it does not establish exact equality or equivalence within a specified margin. Lack of statistical significance is not proof of no effect, and a multiple-testing adjustment does not prove a particular association false. Posterior probabilities condition on the specified estimand, likelihood and priors. A Bayes factor is relative support between specified models, not a posterior probability without prior model odds. A zero shared mean log effect can coexist with positive and negative context effects.

EA6.4. Controls, calibration and shared bias

A negative-control association can signal bias only under an appropriate no-effect assumption for that control. Transferring the control's bias to autism requires an additional comparability assumption. An interval including one neither proves a control unaffected nor certifies an unbiased autism estimate. Bias calibration is therefore conditional. A fake-date profile endpoint is not an empirical upper bound on bias at real events. Independent country errors can average down under the stated covariance model; shared or correlated errors need not. A null-centered precision limit under an assumed shared-bias distribution is not a universal floor on all possible realized intervals.

EA6.5. Interactions, additions and total risk

In the stated two-vaccine mixture, joint risk relative to neither exposure is rA times rB times theta. The effect of adding B to A is rB times theta under the same homogeneous-risk assumptions. Theta equal to one denotes multiplicative separate effects, not total joint risk equal to one; theta above one need not imply total risk above the unexposed baseline. The average interaction across modeled pairs is not a bound on each pair, same-day administration, every subgroup or total schedule load. The largely prior-determined load regressions remain uninformative about their target contrasts.

EA6.6. Growth, risk levels and evidence from overlapping data

Stable measured uptake cannot by itself explain continuing growth under a stable-effect model. That comparison does not establish that exposure has no constant contribution to the outcome level. Growth calculations at assumed individual risk ratios of two and three retain those assumptions; they are not unconditional bounds. Agreement among methods using the same observations assesses sensitivity to those methods, not independent data replication. Shared source errors can affect several analyses.

EA6.7. Selection checks, simulations and numerical agreement

Meeting a recorded selection threshold is not causal validation. Failed pretrends, weak first stages, unfavorable recovery and control results remain reported; excluding a result from a project synthesis does not make its association zero. Low power limits exclusion claims but does not alone invalidate an otherwise identified estimator. Simulation detection and interval-coverage rates describe the specified scenarios and finite repetitions. Agreement between two integration runs is not an error bound or new independent evidence. Current arithmetic agreement does not resolve the historical Swedish standard-error mismatches documented in EA4.

EA6.8. Unresolved event eligibility

Local metadata reproduce 124 selected timing flags and 128 flagged event constituents across 127 bundles. The retained audit reports 42 contributing California configurations whose source outcomes precede the selected policy's recorded date. The exact 42-member disposition and corrected native inputs remain unverified locally. Three further mixed bundles appear in the profile metadata; profile presence is not proof of pool contribution, and their secondary event dates require separate qualification. Earlier potentially valid coevents and all source observations remain retained. These limitations do not authorize substituting another law, changing cohort clocks or silently adopting the later candidate.

EA6.9. Scope of the wider evidence

The wider-evidence conclusion about vaccines and autism is distinct from the certainty of any one public-data model. Individual studies have different exposure contrasts, outcomes, subgroup precision and bias risks. An inverse association does not establish protection, and a null-containing hazard-ratio interval does not negate a positive risk-difference estimate from a different model. Both are retained on their own scales. Recorded diagnoses and service classifications are not direct observations of biological onset.

Methods references: Sauerbrei et al. (2014), STRATOS initiative, Statistics in Medicine 33:5413-5432, doi:10.1002/sim.6265. Wasserstein and Lazar (2016), The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician 70:129-133, doi:10.1080/00031305.2016.1154108. The wider-evidence conclusion follows the WHO GACVS statement of 11 December 2025; it is not inferred from passing this project's checks.