

Table 1 | Prespecified predictor distributions and regression coefficients used to generate datasets A–E.

	Mean	SD	A	B	C	D	E
β_0			5	6	7	8	9
Var_a	1	0.2	1	2	3	4	5
Var_b	2	0.4	6	7	8	9	10
Var_c	3	0.6	3	3	3	3	3
Var_d	4	0.12	6	6	6	6	6
Var_e	5	0.24	5	5	5	5	5
Var_f	6	0.3	−5	−6	−7	−8	−9

Predictor distributions were defined by the specified means and standard deviations (SDs), which were held constant across datasets A–E. The intercept (β_0) and regression coefficients (β) shown for each dataset were used to generate the outcome. Between-study heterogeneity was incorporated by design for the intercept, Var_a, Var_b and Var_f by progressively varying their coefficients across datasets A–E. In contrast, the coefficients for Var_c, Var_d and Var_e were held constant across datasets, representing predictors without intentionally introduced coefficient heterogeneity.

Table 2 | Prespecified correlation matrix for continuous predictors used in dataset generation.

	Var_a	Var_b	Var_c	Var_d	Var_e	Var_f
Var_a	1.00	0.18	0.12	0.10	0.15	0.20
Var_b	0.18	1.00	0.22	0.18	0.32	0.28
Var_c	0.12	0.22	1.00	0.42	0.25	0.36
Var_d	0.10	0.18	0.42	1.00	0.20	0.31
Var_e	0.15	0.32	0.25	0.20	1.00	0.29
Var_f	0.20	0.28	0.36	0.31	0.29	1.00

Pairwise correlations among the six continuous predictors (Var_a–Var_f) were prespecified and applied during generation of the original datasets A–E. The correlation coefficients ranged from 0.10 to 0.42, representing varying degrees of positive correlation among predictors. The same correlation structure was used across all five datasets.

Table 3 | Prespecified regression coefficients for the mixed-predictor simulation datasets A–E.

	A	B	C	D	E
β_0	11	12	13	14	15
Age	1	2	3	4	5
BMI	6	7	8	9	10
Biomarker 1	3	3	3	3	3
Biomarker 2	6	6	6	6	6
Smoking	5	5	5	5	5
Diabetes	−5	−6	−7	−8	−9
Hypertension	−2	−3	−4	−5	−6

Physical activity	-7	-7	-7	-7	-7
CRP	-8	-8	-8	-8	-8

Dataset-specific intercepts and regression coefficients were used to generate the outcome in the mixed continuous and categorical predictor simulation. Between-study heterogeneity was incorporated by design for the intercept, Age, BMI, Diabetes and Hypertension by progressively varying their coefficients across datasets A–E. In contrast, the coefficients for Biomarker 1, Biomarker 2, Smoking, Physical activity and CRP were held constant across datasets, representing predictors without intentionally introduced coefficient heterogeneity.

Table 4 | Prespecified distributions of continuous and categorical predictors used in the mixed-predictor simulation.

Continuous variable	Mean	SD
Age	50	12
BMI	25	4
Biomarker 1	100	20
Biomarker 2	50	10
CRP	10	10
Categorical	No	Yes
Smoking	70%	30%
Diabetes	80%	20%
Hypertension	65%	35%
Physical activity	55%	45%

Continuous variables were generated according to the specified means and standard deviations (SDs), whereas categorical variables were generated as binary predictors according to the prespecified proportions in the “No” and “Yes” categories. These distributions were used consistently across datasets A–E, while between-study heterogeneity was introduced through selected regression coefficients rather than through differences in predictor distributions.

Table 5 | Illustrative study-level inputs used to demonstrate the CoSReMA web application.

Study	n	R ²	Intercept	Var_a β (SE)	Var_b β (SE)	Var_c β (SE)	Var_d β (SE)	Var_e β (SE)	Var_f β (SE)
A	100	0.93	-0.81	0.79 (0.70)	6.20 (0.36)	2.93 (0.29)	5.84 (1.45)	5.58 (0.60)	-4.40 (0.52)
B	100	0.90	3.22	1.62 (0.72)	6.44 (0.34)	2.76 (0.26)	6.61 (1.19)	5.58 (0.65)	-6.03 (0.49)
C	100	0.91	15.09	3.06 (0.88)	8.49 (0.39)	3.83 (0.30)	4.42 (1.39)	3.43 (0.65)	-6.55 (0.55)
D	100	0.89	12.52	5.04 (0.83)	9.47 (0.44)	3.01 (0.31)	5.11 (1.38)	4.83 (0.66)	-8.40 (0.58)
E	100	0.85	14.42	5.70 (1.04)	9.93 (0.52)	3.20 (0.29)	6.38 (1.74)	4.79 (0.92)	-10.20 (0.67)

The table shows the summary statistics entered for five studies (A–E), including sample size (n), reported model R², intercept, and regression coefficients (β) with corresponding standard errors (SE) for predictors Var_a–Var_f. These study-level inputs were used by the application to generate pooled coefficient estimates, predictor-specific forest plots, reconstructed model-level R², and graphical comparisons of study-level and synthesised results.

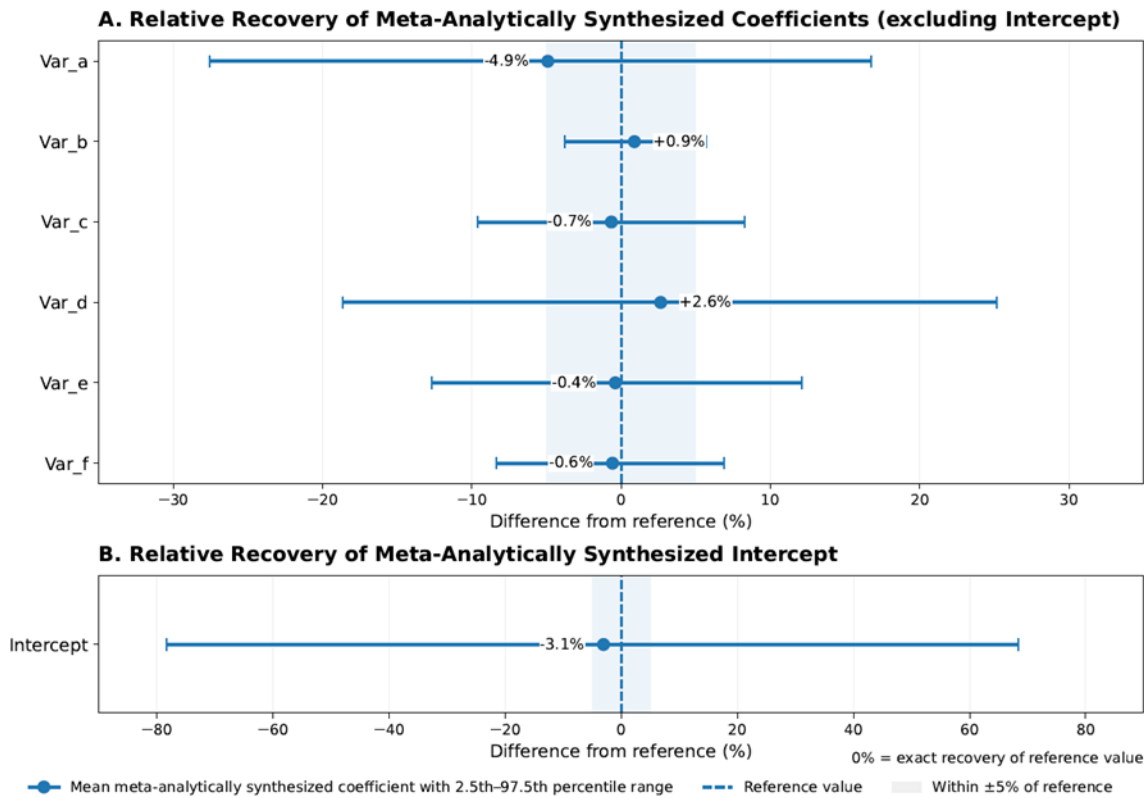
Table 6 | Predictor distribution inputs used in the illustrative CoSReMA web-application analysis.

Parameter	Var_a	Var_b	Var_c	Var_d	Var_e	Var_f
Mean	1.00	2.00	3.00	4.00	5.00	6.00
SD	0.20	0.40	0.60	0.12	0.24	0.30

The table shows the mean and standard deviation (SD) specified for each continuous predictor (Var_a–Var_f). These summary distribution parameters were supplied alongside the study-level regression coefficients, standard errors and R² values and were used by the reconstruction component of CoSReMA to generate study-level synthetic predictor data for estimation of pooled model goodness of fit.

Supplementary Figure 1:

Relative Recovery of Reference Parameters by Meta-Analytical Coefficient Synthesis

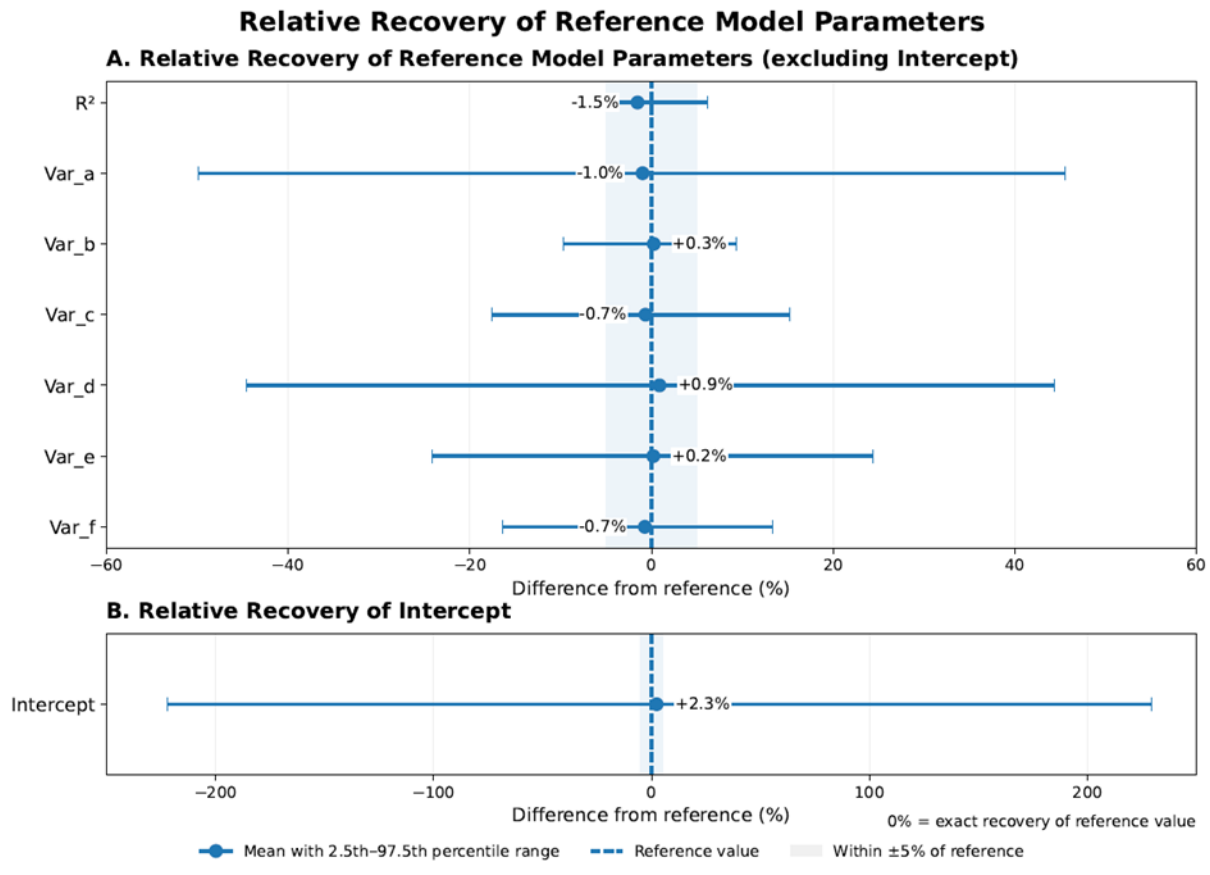


Supplementary Figure 1: Relative recovery of reference regression parameters by meta-analytic coefficient

synthesis. Relative differences between the mean meta-analytically synthesised coefficients across 1,000 simulation iterations and the corresponding coefficients from the pooled individual-level reference model are shown. **a.** Relative recovery of the six predictor coefficients (Var_a–Var_f). **b.** Relative recovery of the intercept, displayed separately because of its wider variability. Points represent the mean pooled coefficient expressed as the percentage difference from the reference value, and horizontal error bars represent the 2.5th–97.5th percentile range across iterations. The dashed vertical line at 0% denotes exact recovery of the reference value, and the shaded region indicates values

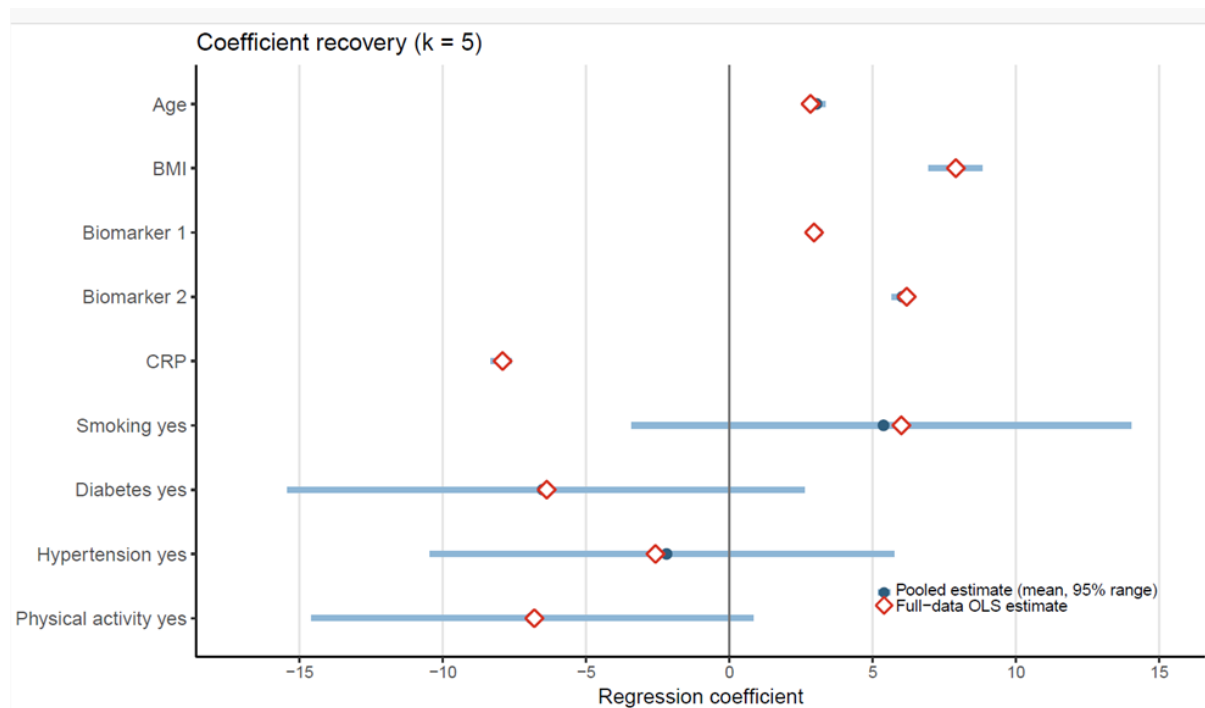
within $\pm 5\%$ of the reference. Percentage labels indicate the relative difference between the mean synthesised estimate and its corresponding reference coefficient.

Supplementary Figure 2:



Supplementary Figure 2: Relative recovery of reference model parameters by summary-statistics-based reconstruction. Relative differences between the mean estimates obtained from reconstructed datasets across 1,000 simulation iterations and the corresponding values from the pooled individual-level reference model are shown. **a**, Relative recovery of R^2 and predictor coefficients (Var_a–Var_f). **b**, Relative recovery of the intercept, displayed separately because of its substantially wider variability. Points represent the mean reconstructed estimate expressed as the percentage difference from the reference value, and horizontal error bars represent the 2.5th–97.5th percentile range across iterations. The dashed vertical line at 0% denotes exact recovery of the reference value, and the shaded region represents values within $\pm 5\%$ of the reference. Percentage labels indicate the relative difference between the mean reconstructed estimate and its corresponding reference value.

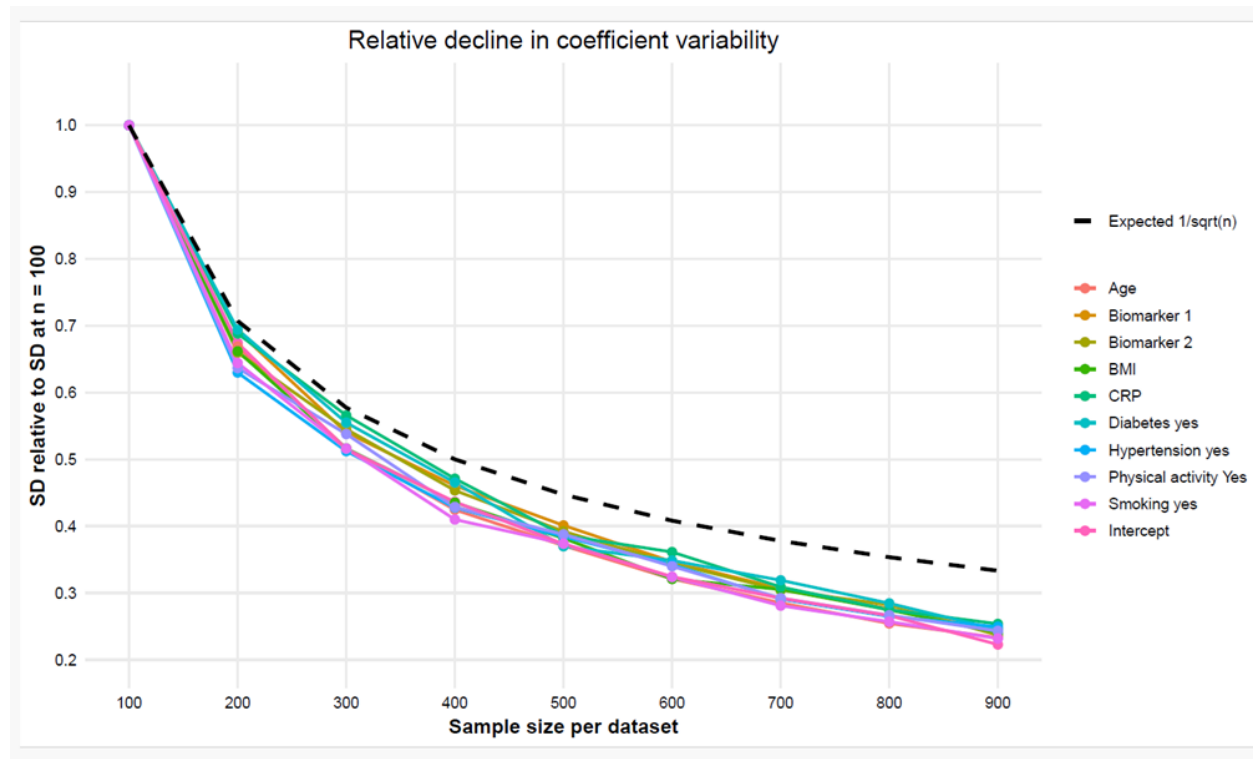
Supplementary Figure 3



Supplementary Figure 3: Recovery of regression coefficients in the mixed continuous and categorical

predictor simulation. Recovery of regression coefficients is shown for the five-study (k=5) simulation containing both continuous (Age, BMI, Biomarker 1, Biomarker 2 and CRP) and categorical (Smoking, Diabetes, Hypertension and Physical activity) predictors. Blue points represent the mean meta-analytically pooled coefficient estimates across simulation iterations, with horizontal bars indicating the corresponding 95% empirical ranges. Red open diamonds indicate the reference coefficients obtained by ordinary least-squares (OLS) regression on the complete pooled individual-level dataset. Close alignment between the pooled and reference estimates indicates coefficient recovery, whereas the width of the empirical ranges reflects variability across iterations. Categorical predictors generally showed greater variability than continuous predictors under the simulated conditions

Supplementary Figure 4



Supplementary Figure 4: Effect of increasing sample size on the variability of recovered regression coefficients.

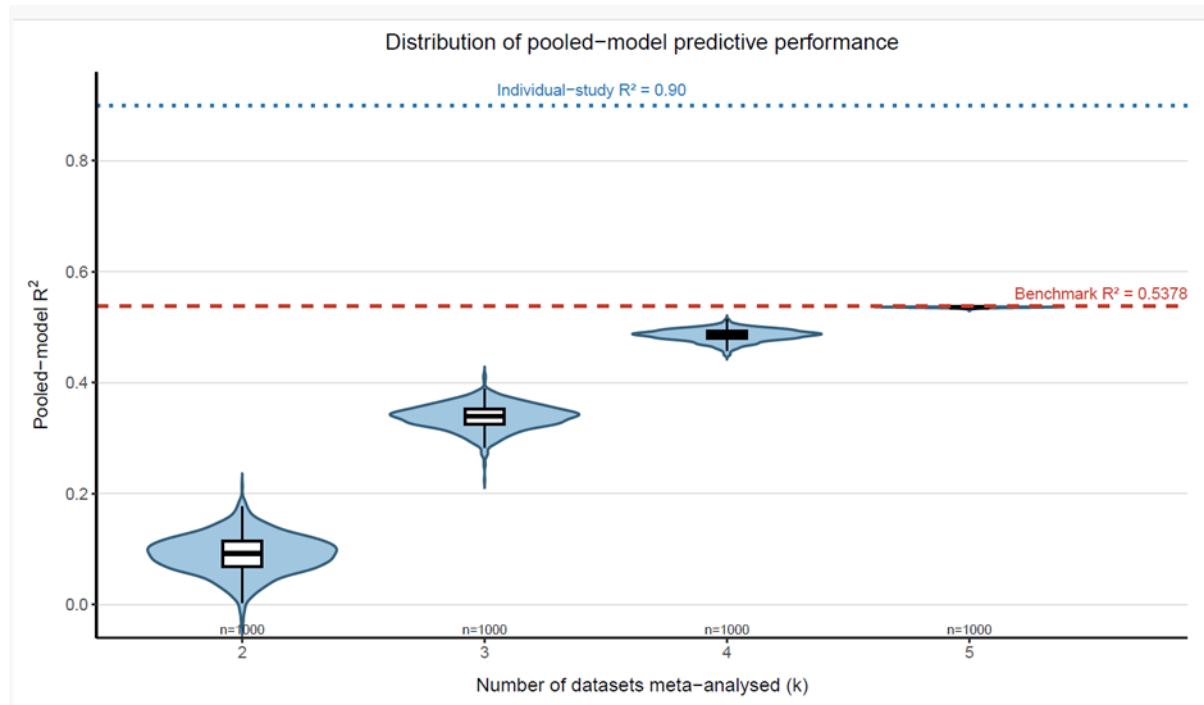
Relative coefficient variability across increasing per-study sample sizes is shown for continuous and categorical predictors and the intercept in the mixed-predictor simulation. For each regression term, variability is expressed as the standard deviation (SD) of the recovered coefficient estimates relative to its SD at a sample size of 100, which was normalised to 1. Increasing the sample size from 100 to 900 observations per dataset produced a progressive decline in coefficient variability across all model terms. The dashed black line represents the expected $1/n$ decline in sampling variability, providing a theoretical reference for the observed reduction in coefficient dispersion with increasing sample size.

Supplementary Figure 5



Supplementary Figure 5: Convergence of reconstructed pooled-model goodness of fit towards the full-data benchmark with increasing number of studies. Mean reconstructed pooled-model R^2 across 1,000 simulation iterations is shown as the number of datasets included in the synthesis (k) increased from two to five. Points represent the mean R^2 , with error bars indicating variability across iterations. The red dashed horizontal line denotes the R^2 of the full pooled individual-level reference model (benchmark $R^2=0.5378$). Percentage labels indicate the proportion of the benchmark R^2 recovered at each value of k . Recovery increased progressively from 16.9% with two datasets to 62.8% with three, 90.2% with four and 99.6% with all five datasets, demonstrating convergence towards the full-data benchmark as additional study populations were incorporated.

Supplementary Figure 6



Supplementary Figure 6: Effect of the number of studies included in synthesis on recovery of pooled-model goodness of fit. Distributions of reconstructed pooled-model R^2 across 1,000 simulation iterations are shown as the number of datasets included in the synthesis (k) increased from two to five. Violin plots show the distribution of R^2 estimates, with embedded box plots indicating the median and interquartile range. The red dashed horizontal line denotes the R^2 of the pooled individual-level reference model (benchmark $R^2=0.5378$), whereas the blue dotted line indicates the R^2 of the individual study models (0.90). Increasing the number of contributing datasets progressively shifted the reconstructed pooled-model R^2 towards the reference benchmark and reduced its variability, with the closest recovery observed when all five datasets were included.

