Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Novel Framework for Meta-analysis of Linear Regression Models

Indrani Devi Sarma
All India Institute of Medical Sciences Guwahati
Dibyajyoti Saikia
dibyajyotisaikia@aiimsguwahati.ac.in

All India Institute of Medical Sciences Guwahati

Research Article

Keywords: Coefficient synthesis, Summary statistics, Data reconstruction, Predictive modelling, Model performance, Evidence synthesis

Posted Date: October 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10958900/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Linear regression models are widely used for prediction and estimation in health research, but model coefficients and performance may vary substantially across populations. Individual-participant-data meta-analysis permits direct synthesis across studies but is often limited by restrictions on data access and sharing. We developed Coefficient Synthesis and Reconstruction Meta-Analysis (CoSReMA), a framework that combines meta-analysis of study-level regression coefficients with reconstruction of synthetic datasets from summary statistics to synthesise both model parameters and overall model fit without requiring participant-level data.

Methods:

A simulation-based methodological study was conducted using five independent datasets of 2,000 observations each, containing six continuous predictors with controlled between-study heterogeneity. The pooled dataset of 10,000 observations provided the reference regression model. Samples of 100 observations were drawn from each dataset, separate multiple linear regression models were fitted, and study-specific coefficients were synthesised using random-effects meta-analysis. For reconstruction, study-level predictor means and standard deviations, regression coefficients, and model coefficient of determination were used to generate 1,000-observation synthetic datasets without imposing the original predictor correlation structure. The reconstructed datasets were combined and refitted. Coefficient and model-fit recovery were assessed across 1,000 iterations, with additional analyses varying sample size, number of studies, and predictor type.

Results:

Meta-analytically synthesised coefficients were centred close to those from the pooled reference model. For example, mean pooled coefficients were 2.90 versus 3.05 for Var_a, 8.01 versus 7.94 for Var_b, and − 6.85 versus − 6.81 for Var_f. Heterogeneity measures appropriately distinguished predictors with deliberately varying effects from those with stable effects. Across 1,000 reconstructions, the mean coefficient of determination was 0.644 (standard deviation 0.021), compared with 0.65 in the pooled reference model. Increasing within-study sample size reduced coefficient variability, while increasing the number of contributing studies improved recovery of pooled model fit.

Conclusions:

CoSReMA recovered complementary components of pooled linear regression models from study-level summary information under the simulated conditions examined. Meta-analysis provided stable pooled coefficient estimates, while summary-statistics-based reconstruction closely reproduced overall model fit. The framework may provide a practical approach for synthesising linear regression models when participant-level data are unavailable, although validation using real-world datasets is required.

Introduction

The use of machine-learning methods in health research has expanded rapidly, with approaches ranging from linear and support vector regression to ensemble algorithms increasingly applied to diagnosis, prognosis and risk prediction[1]. Although such models may achieve high predictive performance in their development datasets, their performance and estimated predictor effects can vary substantially across populations because of differences in demographic characteristics, disease prevalence, clinical practice, measurement procedures and predictor distributions[2].

This heterogeneity poses a fundamental challenge to model generalisability. A model that performs well in one population may not perform similarly in another, making evidence from multiple independent populations essential for assessing the robustness of predictive relationships [3]. Individual-participant-data meta-analysis offers the most direct approach to such synthesis, but access to individual-level data is often restricted by privacy concerns, consent requirements and institutional or regulatory data-sharing policies [4]. Even in prospectively planned multicentre studies, differences in data governance, variable definitions, measurement protocols and population characteristics can make pooling participant-level data difficult[5].

Most published studies also remain accessible primarily through summary statistics, mostly including sample size, predictor distributions, model coefficients and measures of predictive performance but no individual level data[6]. A framework capable of synthesising predictive models from such aggregate information could provide a practical means of evaluating their behaviour and generalisability across populations without requiring access to the original participant-level data[7]. However, no widely adopted and straightforward method currently exists for accomplishing this directly from the summary information typically reported in predictive-modelling studies[8]. Existing meta-analyses in this domain have largely confined themselves to pooling these accuracy metrics, treating the underlying algorithms as opaque objects whose internal parameters are of no synthetic interest[9].

Predictive models are, at their core, collections of estimated parameters, coefficients, weights, splitting rules, and hyperparameters, each of which is fitted to a single dataset and therefore subject to sampling variability, local case-mix, measurement idiosyncrasies, and overfitting[10]. Just as meta-analysis of effect estimates in etiological research yields a pooled value that better approximates the true underlying association than any single study, meta-analytic synthesis of model parameters themselves may yield coefficients that more closely reflect real-world or true population values, rather than the peculiarities of any one development cohort[11]. Extending meta-analytic methodology to the algorithms, their parameters, and their development procedures is thus essential if evidence synthesis is to advance predictive modelling rather than simply audit it[12].

We therefore propose a framework for the meta-analysis of linear regression models using summary-level information. This study aimed to develop and evaluate an approach for synthesising predictive models at the level of their estimated parameters, rather than pooling measures of predictive performance alone.

Meta-analytic synthesis of regression coefficients from study-level estimates has been described previously and provides a means of obtaining pooled estimates of predictor effects[13]. However, coefficient synthesis alone cannot recover the corresponding (R^2). The variance explained by a multivariable regression model depends not only on the regression coefficients, but also on the variances and covariance structure of the predictors and the outcome[14]. Consequently, in the absence of individual-level data of all the studies, the goodness of fit of a model cannot be determined.

A complementary strategy is to reconstruct individual-level datasets from reported summary statistics and estimate model fit from the resulting simulated data. This approach, however, presents a different limitation. Accurate reconstruction of a multivariable dataset requires information on the covariance structure among predictors, whereas predictor correlation matrices are rarely reported[15]. Multicollinearity is well recognised to increase the variance and instability of individual regression coefficients, while overall model fit and fitted values may remain comparatively stable[16]. We therefore hypothesised that model-level goodness of fit may be more robust to incomplete specification of the predictor correlation structure than individual coefficient estimates, and we tested this empirically through simulation.

On this basis, we developed a framework that combines these two complementary sources for evidence synthesis. Regression coefficients are obtained through meta-analytic synthesis of reported study-level estimates, whereas overall model fit is estimated through reconstruction and simulation using the available summary statistics. The framework thus uses coefficient synthesis to estimate model parameters and simulation to recover information on model-level fit, while explicitly recognising the limitations of each approach. We subsequently evaluated whether their combined use could reproduce both the coefficients and R^2 of a reference model derived from the underlying pooled individual-level data. The process is given in Fig. 1.

Accordingly, this study aimed to develop and validate a summary-statistics-based framework for the meta-analytic synthesis of linear regression models. Specifically, we sought to determine whether (i) meta-analytic synthesis of study-level regression coefficients could recover the coefficients of a reference model fitted to the pooled individual-level data; (ii) datasets reconstructed from study-level summary statistics could reproduce the $R^2$of the reference model despite incomplete information on predictor correlations. We further evaluated the robustness of these estimates across repeated simulations and varying sample sizes to determine their bias, variability and convergence towards the reference model. Finally, we aimed to translate the proposed framework into an accessible, user-friendly web application that enables researchers to perform coefficient synthesis, summary-statistics-based data reconstruction and estimation of model-level goodness of fit without requiring individual-participant data or advanced programming expertise.

Methods

Study design and analytical framework

We conducted a simulation-based methodological study to develop and evaluate CoSReMA (Coefficient Synthesis and Reconstruction Meta-Analysis), a framework for reconstructing linear regression models from study-level summary information. The framework combined two complementary components: meta-analytic synthesis of regression coefficients and summary-statistics-based reconstruction of datasets to estimate model-level goodness of fit. Validation assessed whether these approaches could recover the coefficients and R^2 of a reference model fitted directly to pooled individual-level data.

This study did not involve human participants. Clinical trial registration is not applicable.

Generation of reference datasets

Five independent datasets (A–E), each containing 2,000 observations, were generated in R. Six continuous predictors (Var_a–Var_f) were generated from a multivariate normal distribution using prespecified means, standard deviations, regression coefficients (Supplementary Table 1) and correlations (Supplementary Table 2). Dataset-specific regression coefficients were assigned to introduce controlled between-study heterogeneity, and *Change in biomarker* was generated as a linear combination of the predictors with normally distributed residual error calibrated to produce a target R^2 of 0.90.

The five datasets were combined to form a pooled dataset of 10,000 observations. An ordinary least-squares regression model fitted to this complete pooled dataset was designated the reference model, and its coefficients and R^2 were used as benchmarks.

Repeated sampling and coefficient synthesis

To reflect the manner in which evidence is typically generated across independent real-world studies and subsequently synthesized in meta-analysis, a random sample of 100 observations was drawn without replacement from each individual dataset, and separate multiple linear regression models were fitted to each sample. The resulting study-specific regression coefficients were then synthesized meta-analytically and compared with the coefficients obtained from linear regression of the full combined dataset of 10,000 observations, which served as the reference. To assess the robustness and reproducibility of this approach, the entire sampling, regression, and meta-analytic synthesis procedure was repeated over 1,000 iterations. At each iteration, the intercept, regression coefficients, standard errors, confidence intervals, R^2, and predictor means and standard deviations from each study-specific model were retained for subsequent analysis. Pooled coefficients, standard errors, confidence intervals, $\tau 2$, I2, Cochran's Q, and associated P values were recorded.

Summary-statistics-based reconstruction of Datasets

To mimic the conditions of real-world research, in which study-level data are generally observed only once, a single random sample of 100 observations was drawn from each of the five individual datasets. Separate multiple linear regression models were fitted to these samples, and the resulting regression

coefficients, predictor means and standard deviations, and model R^2 values were retained as the only study-level information available for dataset reconstruction. Using these summary statistics, a synthetic dataset of 1,000 observations was reconstructed for each study. Predictors were generated independently from normal distributions using the corresponding study-specific means and standard deviations; the original predictor correlation structure was deliberately not imposed. Residual variance was calibrated to reproduce the study-specific R^2. The five reconstructed datasets were then combined to form a synthetic dataset of 5,000 observations, on which a common multiple linear regression model was fitted. The goodness of fit of the reconstructed model, primarily its R^2, was compared with that of the reference regression model fitted to the original combined dataset of 10,000 observations. To evaluate the stability of the reconstruction procedure, synthetic dataset generation and subsequent regression analysis were repeated over 1,000 iterations while keeping the original n=100 study samples and their derived summary statistics fixed.

Mixed continuous and categorical predictors

A secondary simulation included five continuous and four categorical predictors to evaluate the framework under mixed predictor structures using the same sampling, synthesis, reconstruction and benchmarking procedures (Supplementary Table 3 and 4).

Robustness analyses

The simulation was repeated while varying the number of studies included in synthesis from two to five and the sample size contributed by each study from 100 to 1,000. Recovery of regression coefficients and R^2 was assessed across these scenarios.

Development of the CoSReMA web application

Following validation of the analytical framework, the workflow was implemented as a browser-based web application to enable CoSReMA analyses without requiring specialist statistical programming. Users provide study-level information comprising sample size, R^2, regression coefficients and their standard errors, together with the means and standard deviations or corresponding distributional information for the predictors.

The application performs coefficient synthesis and summary-statistics-based reconstruction locally within the browser using JavaScript. Study-level and pooled estimates are displayed graphically, and analytical outputs can be exported as an Excel workbook with visualisations available in PNG format. An illustrative analysis using five study-level models was performed to verify implementation of the analytical workflow. The application workflow and required input structure are provided in the Supplementary Information.

Results

Reference dataset and pooled benchmark

The five simulated datasets demonstrated consistently high within-study model fit, with R^2 values ranging from 0.89 to 0.90 (Table 1). Despite this similar predictive performance, the underlying regression coefficients differed across datasets for several predictors, reflecting the deliberately introduced between-study heterogeneity. The coefficient for Var_a increased from 1.08 in dataset A to 4.79 in dataset E, while Var_b increased from 6.10 to 10.04. Var_f showed a progressive negative shift, from −4.91 in dataset A to −9.03 in dataset E. In contrast, the coefficients of Var_c, Var_d and Var_e remained comparatively stable across datasets, with values centred around approximately 3, 6 and 5, respectively.

Table 1 | Regression coefficients and model fit for the individual reference datasets and the pooled benchmark model.

	A	B	C	D	E	Combined
R^2	0.90	0.90	0.89	0.90	0.90	0.65
Intercept	5.27	4.28	6.87	7.33	9.84	6.88
Var_a	1.08	2.10	3.08	3.81	4.79	3.05
Var_b	6.10	7.03	8.00	8.94	10.04	7.94
Var_c	2.93	2.99	2.93	3.07	3.06	3.02
Var_d	5.65	6.25	5.90	5.54	5.94	5.68
Var_e	5.10	5.12	4.89	5.18	4.89	5.05
Var_f	−4.91	−5.99	−6.80	−7.72	−9.03	−6.81

Ordinary least-squares regression results are shown for datasets A–E and for the combined individual-level dataset. The individual datasets demonstrated consistently high model fit (R^2 = 0.89–0.90) despite between-dataset variation in several regression coefficients. The combined model was fitted to the pooled individual-level data from datasets A–E and served as the reference benchmark for evaluating coefficient recovery and reconstructed model fit in the CoSReMA framework. R^2, coefficient of determination.

When all five datasets were combined into a pooled dataset of 10,000 observations and a single linear regression model was fitted, the R^2decreased to 0.65. The corresponding pooled-data reference coefficients were 6.88 for the intercept, 3.05 for Var_a, 7.94 for Var_b, 3.02 for Var_c, 5.68 for Var_d, 5.05 for Var_e and −6.81 for Var_f. This pooled model was used as the benchmark against which the performance of coefficient synthesis and summary-statistics-based reconstruction was assessed.

Meta-analytic synthesis of regression coefficients

Random-effects meta-analysis successfully generated pooled estimates for all model terms in each of the 1,000 iterations. Across repeated sampling, the distributions of the pooled coefficients remained centred close to the corresponding reference coefficients derived from the complete pooled dataset (Table 2).

Table 2 | Meta-analytic synthesis of regression coefficients across 1,000 simulation iterations.

	Reference	Mean	Median	SD	2.5th pct	97.5th pct	Mean τ^2	Mean I^2
(Intercept)	6.88	6.67	6.70	2.60	1.49	11.59	11.64	14.63
Var_a	3.05	2.90	2.89	0.35	2.21	3.56	2.10	70.35
Var_b	7.94	8.01	8.01	0.19	7.64	8.39	2.39	92.35
Var_c	3.02	3.00	3.00	0.14	2.73	3.27	0.03	14.81
Var_d	5.68	5.83	5.80	0.64	4.62	7.11	0.57	13.45
Var_e	5.05	5.03	5.02	0.31	4.41	5.66	0.12	13.08
Var_f	−6.81	−6.85	−6.85	0.26	−7.38	−6.34	2.48	86.22

Summary distributions of the random-effects meta-analytically pooled regression coefficients across 1,000 iterations are shown relative to the corresponding coefficients from the pooled individual-level reference model. For each model term, the mean, median, standard deviation (SD), and 2.5th and 97.5th percentiles of the pooled coefficient estimates are reported. Mean between-study variance (τ^2) and mean I^2 across iterations quantify heterogeneity in each regression parameter. Reference denotes the coefficient obtained by fitting the regression model directly to the complete pooled individual-level dataset.

The mean pooled coefficient for Var_a was 2.90 compared with a reference value of 3.05, while the mean for Var_b was 8.01 compared with 7.94. Var_c showed very close recovery, with a mean pooled estimate of 3.00 compared with a reference value of 3.02. Similarly, the pooled estimates for Var_d, Var_e and Var_f were 5.83, 5.03 and −6.85, respectively, compared with reference values of 5.68, 5.05 and −6.81. The corresponding median estimates were also close to the means, indicating that the distributions of repeated pooled estimates were not markedly skewed.

The 2.5th–97.5th percentile ranges were 2.21–3.56 for Var_a, 7.64–8.39 for Var_b, 2.73–3.27 for Var_c, 4.62–7.11 for Var_d, 4.41–5.66 for Var_e and −7.38 to −6.34 for Var_f. The intercept showed substantially greater variability, with a mean pooled estimate of 6.67, compared with the reference value of 6.88, and a 2.5th–97.5th percentile range of 1.49–11.59.

The relative recovery plots showed that the mean pooled estimates of the slope coefficients were generally close to the corresponding reference values (Supplementary Figure 1). The largest average relative deviation among the continuous predictors was observed for Var_a, whereas Var_b, Var_c, Var_e

and Var_f showed only small average deviations. Although the intercept also showed close mean recovery, its much wider percentile range indicated substantially lower precision.

Recovery of between-study heterogeneity

The meta-analytic procedure also reproduced the pattern of heterogeneity deliberately introduced into the simulated datasets. Mean I2 was highest for Var_b at 92.35%, followed by Var_f at 86.22% and Var_a at 70.35%. These were the predictors whose coefficients had been systematically varied across datasets A–E (Figure 2).

In contrast, predictors whose data-generating coefficients were held constant across datasets showed markedly lower heterogeneity. Mean I2 was 14.81% for Var_c, 13.45% for Var_d and 13.08% for Var_e. The mean I2 for the intercept was 14.63%. Thus, the synthesis procedure not only recovered pooled central estimates but also distinguished predictors with substantial between-study variation from those with relatively stable effects.

Recovery of model fit by summary-statistics-based reconstruction

The second component of the framework assessed whether model-level goodness of fit could be recovered from reconstructed datasets generated from study-level means, standard deviations and regression estimates, without imposing the original predictor correlation matrix.

Across 1,000 iterations, the combined reconstructed datasets produced a mean $R^2$of 0.644 with an SD of 0.021. This was close to the reference pooled-data R^2 of 0.65. The median reconstructed R^2 was 0.644, with an interquartile range of 0.630–0.659. The 2.5th and 97.5th percentiles were 0.604 and 0.685, respectively, while the observed minimum and maximum were 0.58 and 0.70. (Table 3)

Table 3 | Recovery of model fit and regression parameters by summary-statistics-based reconstruction across 1,000 simulation iterations.

	Reference	Mean	SD	Median	$P_{2.5}$	$P_{97.5}$
R^2	0.65	0.64	0.02	0.64	0.60	0.69
Intercept	6.88	7.04	7.97	6.67	−8.40	22.67
Var_a	3.05	3.02	0.75	3.03	1.53	4.44
Var_b	7.94	7.96	0.39	7.96	7.17	8.68
Var_c	3.02	3.00	0.25	3.00	2.49	3.48
Var_d	5.68	5.73	1.26	5.74	3.15	8.20
Var_e	5.05	5.06	0.63	5.07	3.83	6.28
Var_f	−6.81	−6.86	0.52	−6.85	−7.93	−5.90

Summary distributions of R^2 and regression coefficients obtained from the combined reconstructed datasets are shown relative to the corresponding values from the pooled individual-level reference model. For each parameter, the mean, standard deviation (SD), median, and 2.5th and 97.5th percentiles across 1,000 reconstruction iterations are reported. Predictor values were reconstructed from study-level summary statistics without imposing the original predictor correlation structure. Reference denotes the R^2 or regression coefficient obtained by fitting the model directly to the complete pooled individual-level dataset.

Adjusted R^2 showed a nearly identical pattern, with a mean of 0.644 and SD of 0.021. The median was 0.644, the interquartile range was 0.630–0.658, and the 2.5th–97.5th percentile interval was 0.603–0.685. The narrow distribution of both R^2 and adjusted R^2 indicated that recovery of overall model fit was relatively stable across repeated simulations.

The mean reconstructed R^2 differed from the reference value by approximately 0.006 in absolute terms. Thus, under the simulation conditions examined, omission of the original predictor correlation structure produced only a small average deviation in model-level goodness of fit.

Variability of coefficients obtained from reconstructed datasets

Although model-level R^2 remained relatively stable, the individual coefficients obtained from the reconstructed datasets showed substantially greater variability across iterations.

The reconstructed intercept had a mean of 7.04 compared with the reference value of 6.88, but its SD was 7.97. Its median was 6.67, with an interquartile range of 1.63–12.48, and the 2.5th–97.5th percentile range extended from −8.40 to 22.67. This contrasted markedly with the narrow distribution observed for R^2. (Supplementary Figure 2)

For the slope coefficients, the mean reconstructed values remained close to the benchmark. Var_a had a mean of 3.02 compared with a reference of 3.05, Var_b 7.96 compared with 7.94, Var_c 3.00 compared with 3.02, Var_d 5.73 compared with 5.68, Var_e 5.06 compared with 5.05, and Var_f −6.86 compared with −6.81. However, the corresponding SDs were 0.75, 0.39, 0.25, 1.26, 0.63 and 0.52, respectively, showing considerably greater between-iteration dispersion than was observed for R^2.

These results indicate that mean coefficient recovery from reconstructed datasets may remain close to the benchmark, but individual estimates are less stable. In contrast, the meta-analytic coefficient estimates were more tightly centred around the corresponding reference values. This difference provides the empirical basis for combining coefficient synthesis with reconstruction-based estimation of goodness of fit within the CoSReMA framework.

Performance with mixed continuous and categorical predictors

The framework was also evaluated in a simulation containing both continuous and categorical predictors. Overall recovery of model-level R^2 remained similar to that observed in the continuous-only simulation. However, the categorical predictors displayed greater coefficient variability than the continuous predictors. (Supplementary Figure 3)

The coefficient-recovery plots showed comparatively narrow distributions for several continuous variables, whereas categorical predictors such as smoking, diabetes, hypertension and physical activity demonstrated substantially wider ranges across repeated analyses. In particular, the confidence or percentile ranges for several binary predictors extended across a broad portion of the coefficient scale.

Thus, the framework remained applicable in a mixed-predictor setting, but coefficient precision was lower for categorical predictors than for continuous predictors under the simulated conditions.

Effect of increasing sample size

Increasing the number of observations sampled from each dataset produced a progressive decline in coefficient variability. The relative SD of the estimated coefficients decreased across the evaluated sample-size range, with all coefficient trajectories showing a monotonic reduction as sample size increased. (Supplementary Figure 4)

The observed decline broadly followed the expected inverse-square-root relationship between sampling variability and sample size. This pattern was evident across the continuous predictors and the intercept, although the rate and magnitude of reduction differed somewhat among model terms. Thus, increasing within-study sample size primarily improved the precision and stability of parameter recovery.

Effect of the number of studies included in synthesis

Model recovery improved as the number of contributing studies increased. The distribution of reconstructed pooled-model R^2 moved progressively towards the full-data benchmark as the number of datasets included in synthesis increased from two to five. (**Supplementary Figure 5 and 6)**

With only two datasets, the reconstructed pooled R^2 remained substantially below the benchmark and showed greater dispersion. Inclusion of a third dataset moved the distribution closer to the full-data estimate, while four datasets produced a further improvement. With all five datasets included, the reconstructed R^2 closely approached the benchmark value shown in the reference model.

This pattern indicates that inclusion of a larger number of heterogeneous study populations improved the ability of the reconstructed pooled model to approximate the relationship observed in the complete pooled dataset.

Implementation of the CoSReMA web application

The analytical framework was subsequently implemented as a browser-based web application. The application accepts study-level summary information including sample size, R^2, regression coefficients, their standard errors, and predictor distribution parameters.

Using the illustrative five-study dataset presented in the manuscript, the application generated study-specific and pooled coefficient estimates, forest plots for individual predictors, reconstructed model-level R^2 estimates, and graphical comparisons of study-level and pooled results (Supplementary Table 5 and 6). The analytical calculations were performed locally within the browser using JavaScript, and outputs could be exported as an Excel workbook, with graphical results available in PNG format. (Figure 3)

The web implementation therefore reproduced the core analytical workflow evaluated in the simulation study while providing an interface that does not require direct use of R or access to the underlying individual-level datasets.

Discussion

This study presents CoSReMA, a framework for synthesising linear regression models from study-level summary information by combining meta-analysis of regression coefficients with reconstruction of synthetic datasets. These complementary approaches address different components of model synthesis, meta-analysis provides pooled parameter estimates, whereas reconstruction enables estimation of model-level goodness of fit. Across the simulations, meta-analytically synthesised coefficients remained close to those obtained from the pooled individual-level dataset, while reconstructed datasets produced R^2 estimates close to the reference pooled model.

An important finding was the contrasting behaviour of regression coefficients and overall model fit. Although individual simulated studies had high R^2 values, several predictor coefficients differed substantially across studies, and the R^2 declined when the heterogeneous populations were combined. This illustrates that strong model performance within individual populations does not necessarily imply equivalent performance in a broader population[17]. Meta-analysis nevertheless recovered the pooled-data coefficients with relatively small average deviations, while heterogeneity statistics appropriately distinguished predictors with varying effects from those with more stable effects. CoSReMA may therefore provide information not only on pooled predictor effects but also on their consistency across populations.

The reconstruction component addresses a limitation of coefficient synthesis: pooled regression coefficients alone cannot determine model R^2 because explained variance also depends on predictor distributions and their covariance structure[16]. Synthetic datasets were therefore reconstructed using study-level predictor distributions, regression estimates, and reported model fit. Despite deliberately omitting the original predictor correlation structure, the reconstructed R^2 remained close to the reference

value, across 1,000 iterations, mean reconstructed R^2 was 0.644 compared with the reference value of 0.65.

This finding should not be interpreted as evidence that predictor correlations are unimportant, as correlations among predictors can influence the precision and stability of individual regression coefficient estimates[16]. Rather, under the conditions examined, overall goodness of fit appeared more robust to incomplete specification of predictor correlations than individual coefficient estimates. Reconstructed coefficients showed considerably greater between-iteration variability, particularly for the intercept and selected predictors, whereas meta-analytically synthesised coefficients were more stable. This contrasting behaviour provides the principal rationale for CoSReMA that regression parameters are obtained through meta-analytic synthesis, while model-level goodness of fit is estimated through summary-statistics-based reconstruction.

The convergence experiments further showed that increasing the number of contributing studies progressively moved reconstructed model fit towards the pooled-data benchmark, while increasing within-study sample size primarily reduced coefficient variability. Secondary simulations incorporating continuous and categorical predictors produced similar overall model-fit recovery, although categorical predictors showed greater coefficient variability. These findings suggest potential applicability beyond continuous-variable models, while indicating that greater uncertainty may occur for categorical predictors, particularly when category prevalence is low or heterogeneous.

A practical advantage of CoSReMA is its reliance on study-level information that may be obtainable from published reports. It is not intended to replace individual-participant-data meta-analysis, which remains preferable when harmonised participant-level data are available because it preserves participant-level information and permits consistent and flexible analyses across studies[18]. Instead, CoSReMA is intended to provide a complementary approach when individual-level data cannot realistically be obtained because of privacy, consent, governance, or other data-sharing constraints[19]. Differences in predictor definitions or measurement procedures across studies may additionally affect the transportability and performance of prediction models and will require careful consideration when applying the framework to real-world evidence [20]. The accompanying browser-based application further translates the framework into an accessible workflow for coefficient synthesis, model reconstruction, and graphical presentation without requiring direct programming.

Several limitations remain. The framework was evaluated primarily under controlled simulation conditions and may behave differently with non-normal predictors, nonlinear relationships, interactions, measurement error, missing data, unequal study sizes, differing predictor definitions, and more complex heterogeneity. Importantly, the simulations do not establish that omission of predictor correlations is generally inconsequential; direct comparisons using different correlation structures are required. The assumption that the same predictors are available across studies also requires further evaluation under realistic patterns of missing or differently coded predictors.

Future work should therefore test CoSReMA across more diverse simulation conditions and, importantly, validate it using real datasets for which both individual-level data and equivalent study-level summary information are available. Such studies would allow direct comparison with conventional individual-participant-data analyses and establish the conditions under which the framework provides reliable estimates.

In summary, CoSReMA demonstrates that complementary components of a pooled linear regression model can be recovered from study-level summary information under the simulated conditions examined. Meta-analysis provided stable pooled coefficient estimates, while summary-statistics-based reconstruction closely reproduced overall model fit. Their combination offers a potential approach to evidence synthesis of multivariable predictive models when individual-level data are unavailable.

Conclusion

CoSReMA provides a methodologically validated framework for the meta-analytic synthesis of linear regression models using study-level summary information. Across repeated simulations, meta-analytic pooling recovered regression coefficients that closely approximated those obtained from the pooled individual-level reference dataset, while summary-statistics-based reconstruction reproduced overall model fit with relatively small average deviation under the conditions examined. The complementary use of these two approaches enables synthesis of both model parameters and goodness of fit without requiring direct access to individual-participant data. These findings indicate that meta-analysis of linear regression models may be feasible using information available in published literature, provided that key summary statistics, including regression coefficients, their uncertainty estimates and predictor distributions, are adequately reported. This also creates opportunities for prospectively designed multicentre studies in which harmonised summary statistics are shared across centres rather than individual-level records, potentially reducing data-transfer, privacy and governance barriers while retaining the ability to synthesise model parameters across populations. Such approaches would not eliminate the need for appropriate data governance or harmonisation, but could substantially reduce dependence on centralised individual-level data sharing. Further evaluation in real-world datasets and under more complex patterns of heterogeneity, missing predictors and correlation structures will be necessary to define the range of settings in which CoSReMA can be applied reliably.

Declarations

Author contributions

D.S. conceived the study, developed the CoSReMA framework, designed the simulation strategy, performed the statistical analyses, interpreted the results, and drafted the manuscript. I.D.S. contributed to study design, methodological development, interpretation of the findings, development of the software and implementation, and critical revision of the manuscript. Both authors reviewed and approved the final manuscript.

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to dibyajyotisaikia@aiimsguwahati.ac.in.

Data availability

All data used in this study were generated through simulation as described in the Methods. The simulated datasets and summary data required to reproduce the analyses are available from the corresponding author upon reasonable request.

Code availability

The R code used for data simulation, coefficient synthesis, dataset reconstruction, statistical analysis and generation of the results is available from the corresponding author upon reasonable request. The CoSReMA web application implements the analytical workflow described in this study.

Funding

No funding was received for this study

Ethics, Consent to Participate, and Consent to Publish declarations are not applicable for this study.

References

1. Rajpurkar P, Chen E, Banerjee O, Topol E. J. AI in health and medicine. Nat Med. 2022;28:31–8.
2. Riley RD, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
3. Debray TPA, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460.
4. Hamilton DG, Hong K, Fraser H, Rowhani-Farid A, Fidler F, Page MJ. Prevalence and predictors of data and code sharing in the medical and health sciences: systematic review with meta-analysis of individual participant data. BMJ. 2023;382:e075767.
5. Neidhart M, et al. A protocol for data harmonization in large cohorts. Nat Ment Health. 2024;2:1134–7.
6. Tudur Smith C, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. Cochrane Database Syst Rev. 2016;9:MR000007.
7. Debray TPA, Koffijberg H, Vergouwe Y, Moons KGM, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. Stat Med. 2012;31:2697–712.
8. Debray TPA, et al. Meta-analysis and aggregation of multiple published prediction models. Stat Med. 2014;33:2341–62.

9. Debray TPA, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019;28:2768–86.
10. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. Biometr J. 2023;65:2200302.
11. Becker BJ, Wu M-J. The synthesis of regression slopes in meta-analysis. Stat Sci. 2007;22:414–29.
12. de Jong VMT, Moons KGM, Eijkemans MJC, Riley RD, Debray TP. A. Developing more generalizable prediction models from pooled studies and large clustered data sets. Stat Med. 2021;40:3533–59.
13. Aloe AM, Becker BJ. An effect size for regression predictors in meta-analysis. J Educ Behav Stat. 2012;37:278–97.
14. Olkin I, Sampson AR. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics. 1998;54:317–22.
15. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. Stat Med. 2011;30:2481–98.
16. Riley RD, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384:e074820.
17. Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol. 2019;72:558–69.
18. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.
19. Ohmann C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open. 2017;7:e018647.
20. Luijken K, Groenwold RHH, van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: a measurement error perspective. Stat Med. 2019;38:3444–59.

Figures

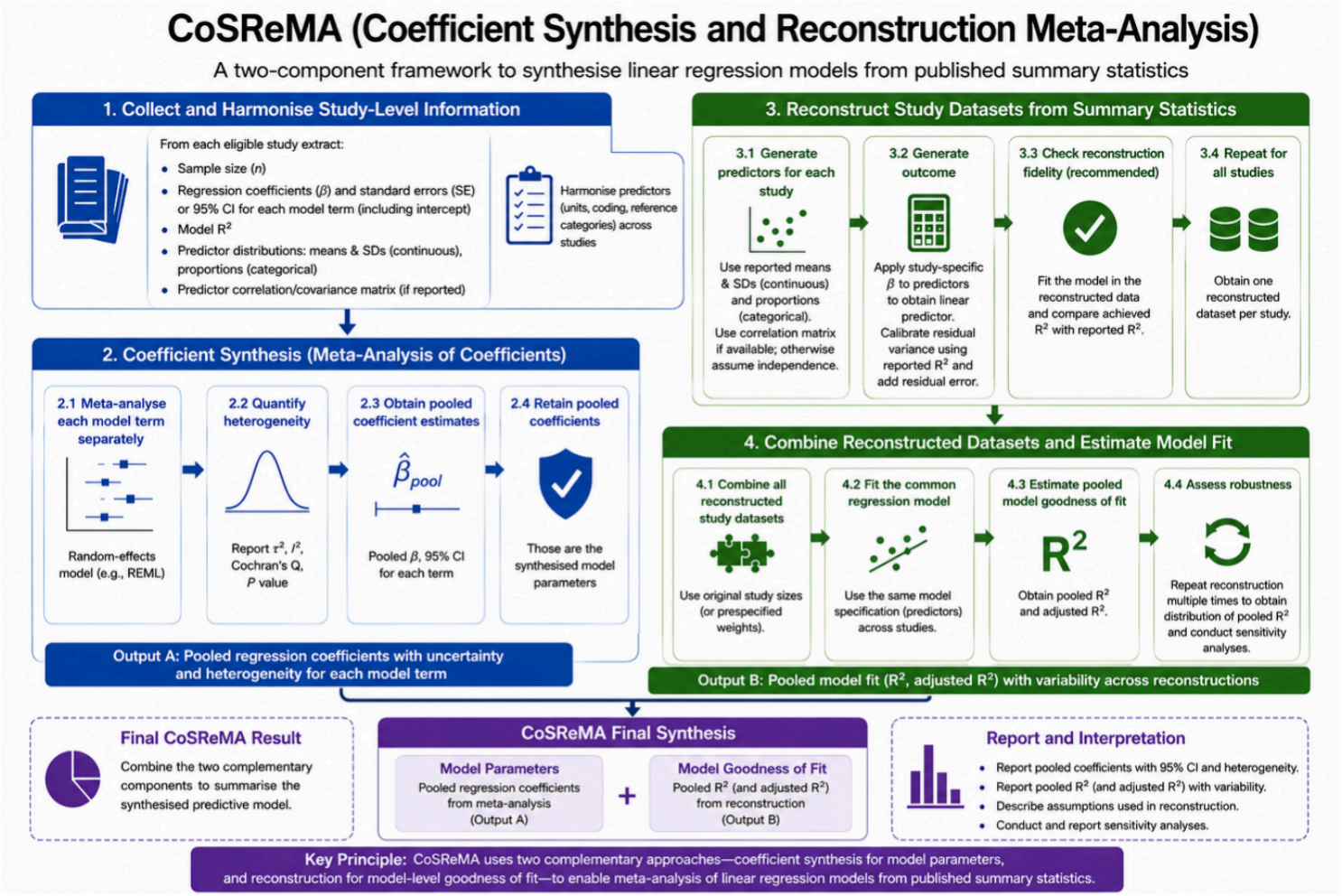


Figure 1

CoSReMA workflow for meta-analysis of linear regression models using summary statistics. The Coefficient Synthesis and Reconstruction Meta-Analysis (CoSReMA) framework integrates two complementary analytical pathways. Study-level regression coefficients are synthesised using meta-analysis to obtain pooled model parameters with measures of uncertainty and between-study heterogeneity. In parallel, predictor distributions, study-specific coefficients and reported R^2 values are used to reconstruct study-level datasets, which are subsequently combined to estimate pooled model goodness of fit. Reconstruction fidelity is assessed by comparing achieved and reported study-level R^2, and repeated reconstruction enables assessment of the variability and robustness of pooled model fit. The final CoSReMA output combines meta-analytically pooled regression coefficients with reconstruction-derived R^2 (and adjusted R^2), together with uncertainty, heterogeneity and sensitivity analyses, enabling synthesis of linear regression models from published summary-level information.

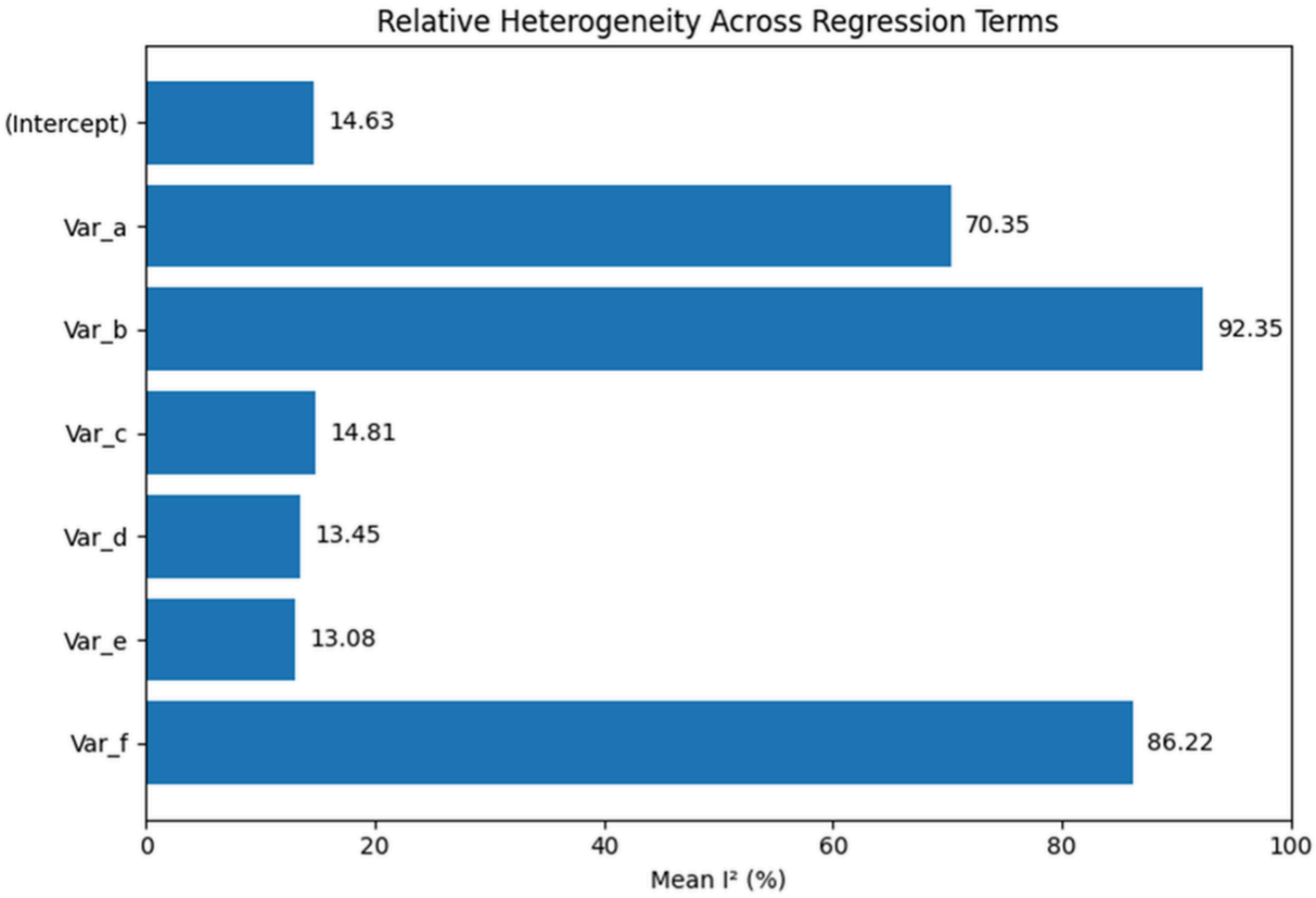


Figure 2

Recovery of between-study heterogeneity across regression terms. Mean I2 values obtained from random-effects meta-analysis across 1,000 simulation iterations are shown for the intercept and each predictor coefficient. Predictors for which between-study heterogeneity was incorporated by design (Var_a, Var_b and Var_f) showed substantially higher mean I2 values than predictors whose coefficients were held constant across datasets (Var_c, Var_d and Var_e). Values shown at the end of each bar represent the mean I2 (%) across iterations.

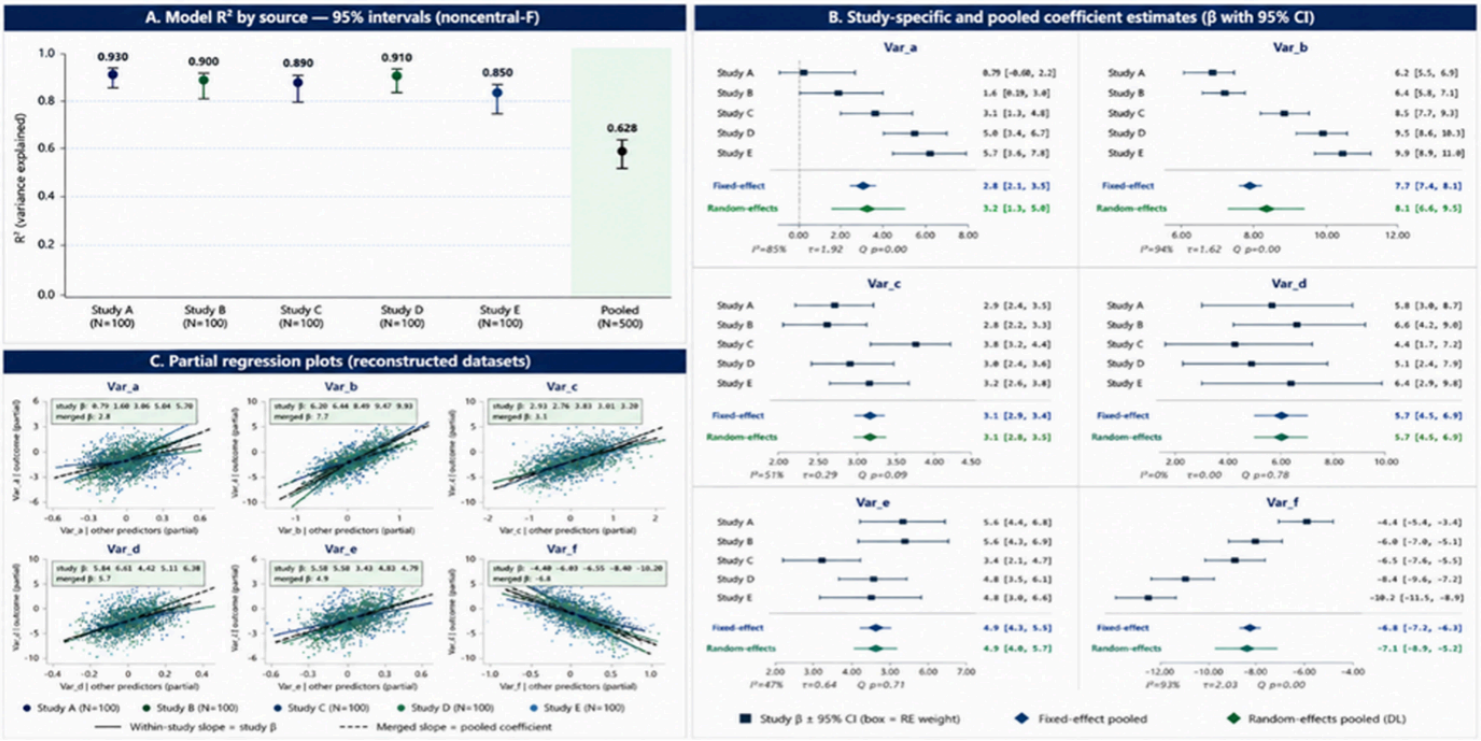


Figure 3

Illustrative outputs generated by the CoSReMA web application from five study-level regression models. a, Study-specific R^2 estimates with 95% confidence intervals and the reconstructed pooled-model R^2, illustrating estimation of overall model goodness of fit from summary-level information. b, Forest-plot representation of study-specific and meta-analytically synthesised regression coefficients with their uncertainty intervals. c, Partial-regression visualisations for predictors Var_a–Var_f in the reconstructed datasets. Points represent reconstructed observations from individual studies; solid lines represent study-specific regression slopes, and the dashed line represents the meta-analytically synthesised coefficient. Together, the panels illustrate the principal outputs of the CoSReMA application: reconstruction-based estimation of pooled model fit, meta-analytic synthesis of regression parameters, and visualisation of study-specific and synthesised predictor–outcome relationships. Link is provided here. genuine-stardust-1d1126.netlify.app

(Username: DSIDS2101

Password @0109meta#2102)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarytablesandfigures.pdf