

METHODOLOGY

Open Access

# Evidence synthesis in prognosis research



Thomas P.A. Debray<sup>1,2\*†</sup> , Valentijn M.T. de Jong<sup>1†</sup>, Karel G.M. Moons<sup>1,2</sup> and Richard D. Riley<sup>3</sup>

## Abstract

Over the past few years, evidence synthesis has become essential to investigate and improve the generalizability of medical research findings. This strategy often involves a meta-analysis to formally summarize quantities of interest, such as relative treatment effect estimates. The use of meta-analysis methods is, however, less straightforward in prognosis research because substantial variation exists in research objectives, analysis methods and the level of reported evidence.

We present a gentle overview of statistical methods that can be used to summarize data of prognostic factor and prognostic model studies. We discuss how aggregate data, individual participant data, or a combination thereof can be combined through meta-analysis methods. Recent examples are provided throughout to illustrate the various methods.

**Keywords:** Prediction, Meta-analysis, Prognosis, Validation, IPD

## Content

Thorough and systematic appraisal of the existing evidence has become mainstream in medical research and practice [1, 2]. Over the past few decades, meta-analysis has been put forward as the de facto statistical method for summarizing the results from a systematic review and appraisal of existing data on a certain topic. In meta-analysis, estimates of interest (e.g., for a specific treatment effect [3] or diagnostic test-outcome association) are obtained from individual studies and then combined into a weighted average. Such quantitative data synthesis potentially increases statistical power to detect genuine associations or effects, to investigate sources of variation within and across studies, and to answer questions that were not posed by individual studies [4, 5].

Meta-analysis is commonly applied in the domain of randomized therapeutic intervention studies [3] and, more recently, in that of diagnostic test accuracy studies.

In the current era of personalized or precision medicine, the use of prognostic information is considered increasingly important to predict outcomes of individuals (off or on treatment) in order to make tailored treatment decisions [6–11]. It therefore seems timely to apply meta-analytic approaches that allow the quantitative synthesis of prognostic evidence [12].

Key barriers of quantitative synthesis of data from prognosis studies are, among others, the lack of high-quality data often due to poor reporting, lack of uniformity in statistical analysis across studies, lack of agreement on relevant statistical measures, and lack of meta-analytical guidance for synthesis of prognosis study data. Recently, much guidance has been written on how to define a review question [13], define the PICOTS (Patients, Index prognostic factor or model, Comparator factor or model, Outcomes, Timing of prognostication, Setting of prognostication), define the search strategy, design the data extraction list [14], and do risk of bias assessments [14, 15]. However, there is relatively little guidance on how to do the actual meta-analysis of results from prognosis studies.

In this paper, we discuss how the data or prognostic results from individual studies, routine care sources (e.g., hospital records or registries), and biobanks can

\*Correspondence: [T.Debray@umcutrecht.nl](mailto:T.Debray@umcutrecht.nl)

†Thomas P.A. Debray and Valentijn M.T. de Jong contributed equally to this work.

<sup>1</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

<sup>2</sup>Cochrane Netherlands, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

Full list of author information is available at the end of the article



be combined quantitatively. Hereto, we describe statistical methods for the meta-analysis of aggregate data (AD), individual participant data (IPD), or a combination thereof. The aim of this gentle overview is to inform researchers of available methods for synthesis of data of prognostic factor and prognostic model studies and to encourage their use when individual studies fail to provide generalizable evidence, as we wish to highlight recent advances in these fields.

### Quantitative synthesis in prognostic factor research

Estimates of overall prognosis (e.g., population outcome risk) are rarely sufficient to inform treatment recommendations and individual patient management. For this reason, it is often helpful to distinguish groups of people with a different average prognosis [6, 7]. A common approach is to identify specific factors that, among people with a given startpoint (such as diagnosis of disease), are associated with a subsequent endpoint [8]. This generally requires estimation of a factor-outcome association which can, for instance, be quantified using a hazard ratio or an odds ratio [8].

Several meta-analysis methods can be used to generate summary estimates of the association between a prognostic factor and a certain outcome. Although it is fairly straightforward to summarize crude (i.e., unadjusted) estimates of a particular factor-outcome association, this practice is generally discouraged because in practice hardly any prognostication is done based on a single factor only [16, 17]. For this reason, we here focus on meta-analysis methods to summarize the adjusted estimates of a certain prognostic factor and outcome. An overview of the presented methods is provided in Table 1.

### Meta-analysis of prognostic factor estimates using aggregate data

A relatively simple situation arises when the prognostic factor of interest is unadjusted in all studies, or has been adjusted for the same other prognostic factors (covariates) in all studies. Traditional meta-analysis methods—as used in meta-analysis of intervention studies—can then be used to summarize the corresponding aggregate data (AD) [18]. The most well-known approach, also from other types of meta analysis, is the so-called fixed effect meta-analysis approach, which can be formulated as follows [19, 20]:

$$\hat{\theta}_i \sim \mathcal{N}(\mu, \hat{s}_i^2) \quad (1)$$

where  $\hat{\theta}_i$  is the estimated factor-outcome association (e.g., log hazard ratio) from the  $i^{\text{th}}$  study, with an estimated standard error  $\hat{s}_i$ . This approach yields a summary estimate of the prognostic effect ( $\mu$ ), which simply represents a weighted average of the  $\hat{\theta}_i$ s.

A common interpretation of fixed effect meta-analysis is that the *true* factor-outcome association is identical for all studies (i.e.,  $\theta_i = \mu$ ). In practice, however, true values for factor-outcome associations are likely to vary across studies due to differences in, e.g., study design, follow-up, variable definitions, adjustment factors, settings, and healthcare standards. It may therefore be more reasonable to assume that the factor-outcome associations  $\theta_i$  are unrelated and to adopt a fixed effects meta-analysis [21]. In this approach, the weight for each study is proportional to both the number of study participants and to how much information is contributed per subject. The meta-analysis then produces an average effect applicable to an amalgamation of the contributing study populations.

**Table 1** Available methods for quantitative synthesis in prognostic factor research

	Available data	Estimate of interest	Possible methods for evidence synthesis
AD	Baseline characteristics	Linear FOA	Meta-regression [34]
	Similarly adjusted FOAs	Linear FOA	Univariate meta-analysis [19], multivariate meta-analysis [20, 33, 57]
		Non-linear FOA	Univariate meta-analysis [36, 37], multivariate meta-analysis [35]
	Not similarly adjusted FOAs	Linear FOA	Multivariate meta-analysis [33, 56, 57]
IPD		Linear FOA	One-stage meta-analysis [34, 38], two-stage meta-analysis [38], multivariate meta-analysis [38, 56], graphical meta-analysis [98]
		Non-linear FOA	One-stage meta-analysis [34, 41], two-stage meta-analysis [37, 41], multivariate meta-analysis [43]
IPD + AD	Baseline characteristics	Linear FOA	Hierarchical-related regression [34]
		Non-linear FOA	Hierarchical-related regression [34]
	Similarly adjusted FOAs	Linear FOA	Two-stage meta-analysis, hierarchical-related regression [49]
		Non-linear FOA	Two-stage meta-analysis [37], hierarchical-related regression [34]
	Not similarly adjusted FOAs	Linear FOA	Multivariate meta-analysis [51, 56], adaptation method [53, 54]

FOA factor-outcome association

Finally, a third option is to adopt a so-called random effects meta-analysis approach, which assumes that the factor-outcome associations  $\theta_i$  are different but related across studies. A major advantage of this approach is that the presence of between-study heterogeneity can directly be quantified [19, 20]:

$$\hat{\theta}_i \sim \mathcal{N}(\mu, \tau^2 + \hat{s}_i^2) \quad (2)$$

The random effects model includes an additional parameter  $\tau$  representing the (unknown) between-study standard deviation. The overall summary result ( $\mu$ ) now represents the average (mean) prognostic effect of the factor across the studies.

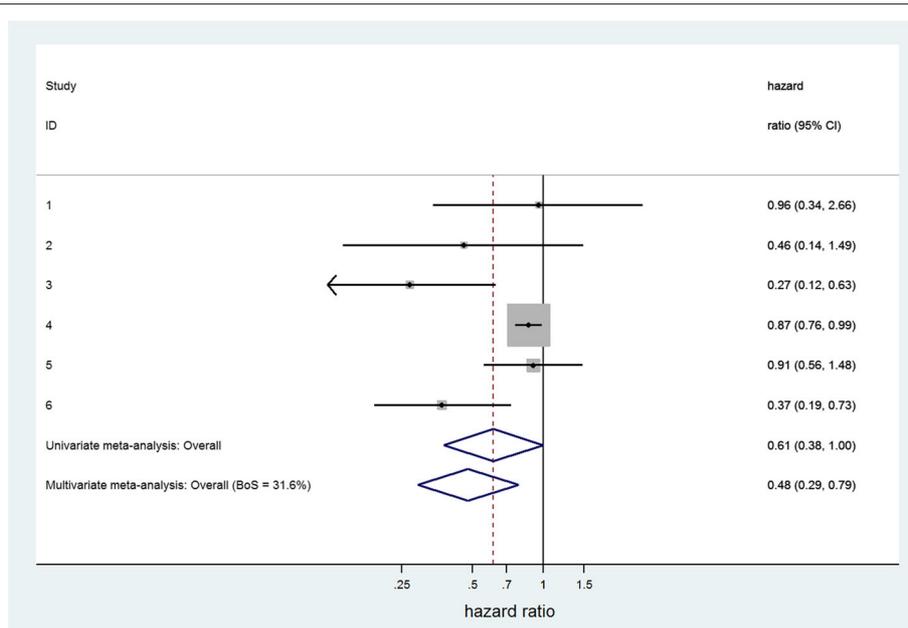
Several methods exist for estimating the weighted average  $\mu$  and the between-study standard deviation  $\tau$  [22, 23]. One approach is to estimate  $\mu$  and  $\tau$  simultaneously, e.g., by adopting (restricted) maximum likelihood estimation. Alternatively, it is possible to first estimate  $\tau$  and then use the corresponding value to obtain an estimate for  $\mu$ . When this strategy does not take the uncertainty of  $\tau$  into account, confidence intervals for  $\mu$  may become too narrow [24]. For this reason, it is generally recommended to adjust these intervals using the methods proposed by Hartung and Knapp [25] and Sidik and Jonkman [26].

As an example, Zhang et al. previously investigated the prognostic effect of progesterone receptor status

in cancer-specific survival in endometrial cancer [27]. Aggregate data from 6 studies were pooled using a random effects meta-analysis (Der Simonian and Laird method), yielding a summary hazard ratio of 0.62 and a corresponding 95% confidence interval (95% CI) ranging from 0.42 to 0.93. When adopting restricted maximum likelihood estimation, the summary estimate changed to 0.61 with a 95% CI from 0.38 to 1.00 (Fig. 1). The wider CI is due to a larger estimate of  $\tau$  when using restricted maximum likelihood estimation rather than DerSimonian and Laird.

### Multivariate meta-analysis

Whereas traditional meta-analysis methods are applied to summarize multiple estimates of a single parameter, it is also possible to jointly summarize multiple estimates of two (or more) parameters using so-called bivariate (or multivariate) meta-analysis methods [20, 28, 29]. These methods are well known in the meta-analysis of diagnostic test accuracy, where one jointly estimates the sensitivity and specificity of the test under review [30]. Multivariate meta-analysis methods aim to account for the correlation between the different parameter estimates and can therefore be used to deal with situations where two or more correlated parameters/statistics are to be synthesized per study. The (bivariate) random effects model for jointly summarizing the AD for two parameters of interest is given as follows:



**Fig. 1** Forest plot for prognostic effect of progesterone on cancer specific survival in endometrial cancer, with summary results for univariate and multivariate meta-analysis. The multivariate meta-analysis of cancer specific survival and progression-free survival used the approach of Riley et al. to handle missing within study correlations, through restricted maximum likelihood estimation [33]. Heterogeneity was similar in both univariate and multivariate meta-analyses ( $I^2 = 70\%$ )

$$\begin{pmatrix} \hat{\theta}_{1i} \\ \hat{\theta}_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} + \begin{pmatrix} \hat{s}_{i1}^2 & \hat{r}_i\hat{s}_{i1}\hat{s}_{i2} \\ \hat{r}_i\hat{s}_{i1}\hat{s}_{i2} & \hat{s}_{i2}^2 \end{pmatrix}\right) \quad (3)$$

where  $\hat{r}_i$  and  $\rho$  represent the (estimated) within-study and, respectively, the (unknown) between-study correlation coefficients. For example,  $\hat{\theta}_1$  and  $\hat{\theta}_2$  may be the prognostic effect on outcome 1 and outcome 2, respectively.

A common application of multivariate meta-analysis arises when researchers are interested in a prognostic factor's association with multiple outcomes [28]. For instance, in the endometrial cancer example, the unadjusted hazard ratio (HR) of progesterone was estimated for cancer-specific survival (6 studies) and for progression-free survival (11 studies). The corresponding hazard ratios of the 17 studies were then jointly pooled using a bivariate random effects meta-analysis [28]. As illustrated in Fig. 1, this strategy yielded a different and more precise summary estimate of cancer-specific survival (unadjusted HR = 0.48, 95% CI 0.29 to 0.79) as compared to the univariate meta-analysis approach above (unadjusted HR = 0.61, 95% CI 0.38 to 1.00).

Multivariate meta-analysis can also be used to jointly summarize prognostic factor-outcome associations that have been adjusted for different sets of prognostic factors (covariates). Researchers then need to distinguish between estimates that are adjusted for all relevant covariates, and estimates that are only adjusted for some (but not all) of the relevant covariates.

Unfortunately, the within-study correlations  $\hat{r}_i$  are rarely reported, thereby complicating the multivariate meta-analysis approach. Riley previously demonstrated that simply ignoring these correlations can lead to meta-analysis results with inferior statistical properties [31]. Researchers may therefore assume a common within-study correlation (e.g.,  $\hat{r}_i = 0$  for all studies), recover its magnitude from reported summary statistics [32], or replace all within- and between-study correlations by an

overall correlation parameter that is estimated from the AD at hand [33].

**Other meta-analysis approaches**

Several extensions for AD meta-analysis of prognostic factor studies have been proposed and can be used to explore sources of between-study heterogeneity [20, 34], to combine studies with different methods of measurement [35], or to combine studies that categorized continuous factors [35–37].

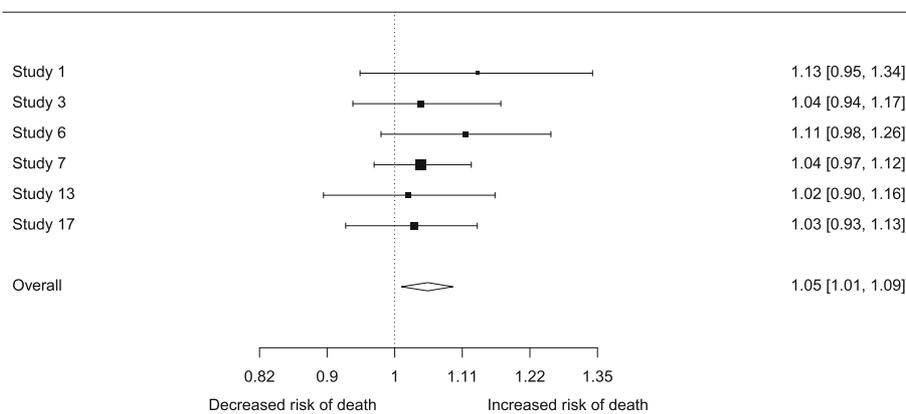
**Meta-analysis using individual participant data (IPD)**

When IPD are available from multiple prognostic factor studies, various random effects meta-analysis models are possible that employ a one-stage or two-stage approach [3, 38, 39].

**Two-stage meta-analysis**

In the two-stage approach, each study is first summarized by its factor-outcome association estimate and standard error. These AD are then appropriately combined across studies into a summary effect using traditional meta-analysis methods. For instance, Trivella et al. performed a two-stage IPD-MA to investigate the role of angiogenesis as a prognostic factor in patients with non-small-cell lung carcinoma [40]. They estimated the log hazard ratio of microvessel-density counts for each participating study center, adjusted for age and cancer stage. These estimates were then pooled using random effects inverse-variance meta-analysis (Fig. 2).

The two-stage IPD-MA approach can also be used to summarize the association of non-linear prognostic factors [41, 42]. In the first stage, the factor-outcome association of interest is modeled separately for each study with a certain functional form (e.g., cubic spline) and parameterization (e.g., location of knots). An overall function can then be obtained in the second stage by meta-analyzing



**Fig. 2** Meta-analysis of multivariable predictor effects. Association between risk of death and increase of one microvessel count, as measured by the Chalkley method. Estimates represent multivariable hazard ratios, adjusted for age and cancer stage [40]

the study-specific function values for distinct factor values [41, 42].

For instance, Sauerbrei et al. combined IPD from nine population-based registries to study the prognostic effect of age in T1-2 breast cancer patients [41]. They estimated a Cox regression model separately in each registry, and adjusted for 5 to 10 other prognostic factors such as the type of surgery and radiotherapy. Studywise selected fractional polynomials (FP) were used to model the adjusted effect of age. The resulting FP functions were then averaged pointwise, with weights for each registry depending on the variance of the the log relative hazard at distinct age values. Results indicated that the mortality risk is low for women between about 40 and 65 years, and increases outside this range.

#### **Multivariate (two-stage) meta-analysis**

Also for IPD meta-analysis, it is possible to simultaneously analyze multiple outcomes by adopting multivariate meta-analysis methods. This typically involves a two-stage approach where the IPD of each study is first reduced to AD (including estimates of the within-study correlation) and subsequently pooled across studies. Multivariate meta-analysis methods have, for instance, been proposed to summarize the association of (non-linear) continuous markers [43]. In the first stage, a common function (e.g., spline with a common location and number of knots for all studies) is estimated separately in each study. The resulting AD (e.g., multivariable regression coefficients) are then pooled across studies in the second stage. In contrast to univariate pooling of estimated effects on a grid of exposure values [41], a major advantage of this approach is that it better accounts for correlations, thereby decreasing bias and improving precision.

#### **One-stage meta-analysis**

An alternative approach for IPD meta-analysis (IPD-MA) of prognostic factor studies is a one-stage approach which synthesizes the IPD from all studies in a single step, while accounting for clustering of patients within studies [44, 45]. The estimation of a pooled factor-outcome association then involves the fitting of a mixed effect model, where each parameter (e.g., regression coefficient) can be specified as common, random or independent (fixed) across studies. One-stage methods appear particularly advantageous when few studies or few patients per study are available [38], or when studies involve time-to-event outcomes [46, 47].

For instance, Den Ruijter et al. performed a one-stage meta-analysis using IPD from 14 cohorts to estimate the association between (log-transformed) carotid intima-media thickness (CIMT) and the incidence of first-time myocardial infarction or stroke [48]. They first assessed between-study heterogeneity by estimating statistical

interaction between cohort and CIMT measurements. Subsequently, a multivariable Cox proportional-hazards model was fitted with random effects for the baseline hazard and common effects for the regression coefficients.

When adopting a one-stage approach, it is generally recommended to account for potential ecological bias [34]. This bias may, for instance, arise when patient outcomes are associated with the mean value of the prognostic factor, rather than the individual covariate values. Ecological bias can be mitigated by separating the within-study and across-study associations, as described elsewhere [49].

#### **Meta-analysis using IPD and AD**

Although IPD meta-analyses are generally considered as the gold standard, IPD cannot always be obtained from all relevant studies. To avoid (data availability) bias, it is often helpful to supplement the available IPD with AD for those studies where IPD are not available [50]. This strategy can be implemented using the approaches described below, assuming suitable AD can be obtained from the non-IPD studies.

#### **Two-stage meta-analysis**

A simple approach is to generate AD from each available IPD set and to jointly summarize the newly derived (from IPD studies) and previously published AD (from non-IPD studies) using aforementioned meta-analysis methods for AD [50]. When critical information from the non-IPD studies is missing (e.g., within-study correlations), the IPD studies can be used to derive the relevant statistics, thereby reducing the risk of bias in summary estimates [31, 35, 51, 52].

A specific situation arises when the non-IPD studies provide factor-outcome associations that are not adjusted for all relevant covariates. A two-stage bivariate meta-analysis can then be used to combine these partially adjusted estimates with the (fully and partially adjusted) factor-outcome associations from the IPD studies.

#### **The adaptation method**

As mentioned earlier, it is common that AD studies do not adjust for all relevant covariates and only provide factor-outcome associations that are partially adjusted. An alternative method to combine fully adjusted associations with the partially adjusted ones is to use the difference in value between the corresponding regression coefficient(s) [53, 54]. This difference is first estimated in the IPD at hand, and then applied to the summary estimate of the partially adjusted factor-outcome association. The adaptation method has, for instance, been applied in a study investigating risk factors for methicillin-resistant *Staphylococcus aureus* acute bacterial skin and skin structure infections [55]. The study authors conducted a literature review to retrieve unadjusted odds ratios for 7 potential

risk factors. These odds ratios were then summarized for each risk factor using a random effects meta-analysis and adapted into an adjusted odds ratio using the IPD at hand.

The adaptation method is strongly related, and in some situations even equivalent, to the aforementioned two-stage meta-analysis approach [56]. Although formal comparisons are lacking, it has been argued that the adaptation method may be less statistically and computationally efficient.

#### **Hierarchical-related regression**

This one-stage approach directly combines the available IPD and AD by specifying a distinct likelihood for each data source [34, 49]. This enables the IPD studies to contribute in all parameter estimates, whereas the AD studies are only used to estimate the study-level parameters and across-study relationships. For example, Riley and Steyerberg adopted hierarchical-related regression to investigate the relationship between age and the risk of 6-month mortality in patients with traumatic brain injury (TBI) [34]. They used a Bernoulli distribution to model the binary outcomes from 4 IPD studies and a Binomial distribution for the observed event counts in 10 AD studies. To account for potential ecological bias, the within-study and across-study effects for participant age were separated when jointly analyzing the 14 studies. It was found that an individual's probability of death by 6 months increases as their individual age increases and also as the mean age in their study (or population) increases. A possible explanation for this is that studies with a higher mean age involved clinicians with less experience of treating TBI patients.

#### **Summary points**

Evidence synthesis in prognostic factor research may help to identify factors that are associated with a certain clinical outcome, to explore their functional form, and to quantify their incremental value over established prognostic factors [8]. When IPD are unavailable, traditional meta-analysis methods can be used to summarize published prognostic factor estimates in order to identify genuine prognostic factors [18]. Although IPD are not strictly required to assess the incremental value of a prognostic factor or to explore its functional form, this may often be unfeasible using published AD only [44]. For this reason, when IPD are available for a few studies, corresponding information can be used to restore unreported AD (e.g., missing within-study correlations) or to adapt unadjusted factor-outcome associations. Evidence synthesis in prognostic factor research is, however, most appealing when multiple sources of IPD are available, as this allows to derive desired prognostic factor results directly and to analyze continuous factors more appropriately [8]. Meta-analysis of IPD is preferably initiated using a two-stage approach, as corresponding methods

are relatively straightforward to implement and guard against ecological bias. One-stage meta-analysis methods may, however, be more appealing when few studies or few subjects per study are available, as they are more flexible, resistant against small sample bias, and avoid the need for estimating correlations between random effects [38].

#### **Quantitative synthesis in prognostic model research**

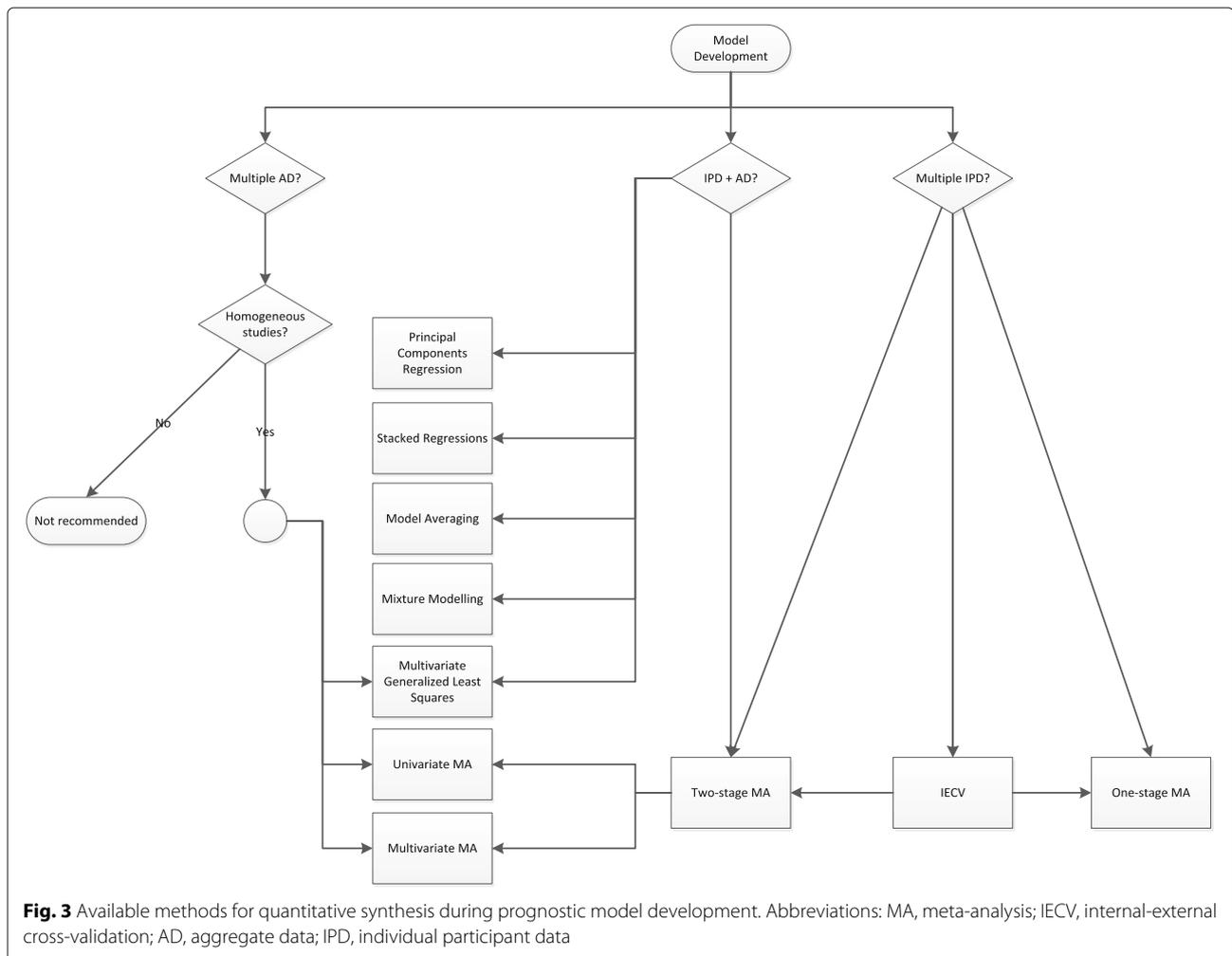
Prognostic model research aims to examine multiple prognostic factors in combination [6], in order to predict the absolute risk of future outcomes in single individuals. Corresponding studies may derive new prognostic models (so-called development studies), evaluate the performance of existing models in new individuals (so-called validation studies) and if necessary tailor their predictions, or examine the model's impact on health-related outcomes.

Currently, most prognostic models are developed based on relatively small studies. Hence, many of these models do not perform adequately when applied to other individuals [9, 58–60]. To investigate and improve the performance of prognostic models across different settings and populations, researchers may consider meta-analysis methods during their development and validation [6, 61–65]. Several strategies for this purpose are described below and summarized in Figs. 3 and 4. As before, we distinguish between situations where the available data sources comprise of aggregate data, individual participant data, or a combination of both.

#### **Meta-analysis using AD**

##### **Validation of an existing prognostic model**

A common source of AD are so-called external validation studies assessing the (discrimination and calibration) performance of a certain prognostic model when tested in other individuals than from which the model was developed. By summarizing these performance estimates, it becomes possible to identify whether the model's predictions are sufficiently accurate across different settings and populations. This typically requires the retrieval of multiple performance statistics (e.g., concordance statistic, calibration-in-the-large, calibration slope) and corresponding standard errors [66, 67]. The resulting estimates can then be pooled using traditional meta-analysis methods, provided that an appropriate scale [68] or link function [67, 69] is used. Although different study weights can be used [21, 70], it is generally recommended to allow for between-study heterogeneity as validation studies are likely to differ in their design and execution [66–68]. As is the case in meta-analysis of prognostic factor research, meta-regression can be used to explore potential sources of between-study heterogeneity.



For instance, van Doorn et al. reviewed 19 published validations of CHA<sub>2</sub>DS<sub>2</sub>-VASc, a prediction model for estimating stroke risk in patients with atrial fibrillation [71]. A random effects meta-analysis was applied to summarize estimates of model discrimination (logit c-statistic) and annual risk per score (square root risks). The summary c-statistic was 0.64 (95% CI 0.56–0.71), which increased to 0.71 (95% CI 0.62–0.79) for studies recruiting patients from a hospital care setting. Further, stroke risks were found to vary substantially within the different scores and were notably elevated in hospital patients as compared to patients from the general population.

#### **Development of a new prognostic model**

It is also possible to summarize AD from multiple but similar prognostic model development studies and to combine their regression coefficients into a new prediction model (for example, via a multivariate meta-analysis) [32, 57]. This strategy is, however, often complicated by the poor reporting of key model parameters

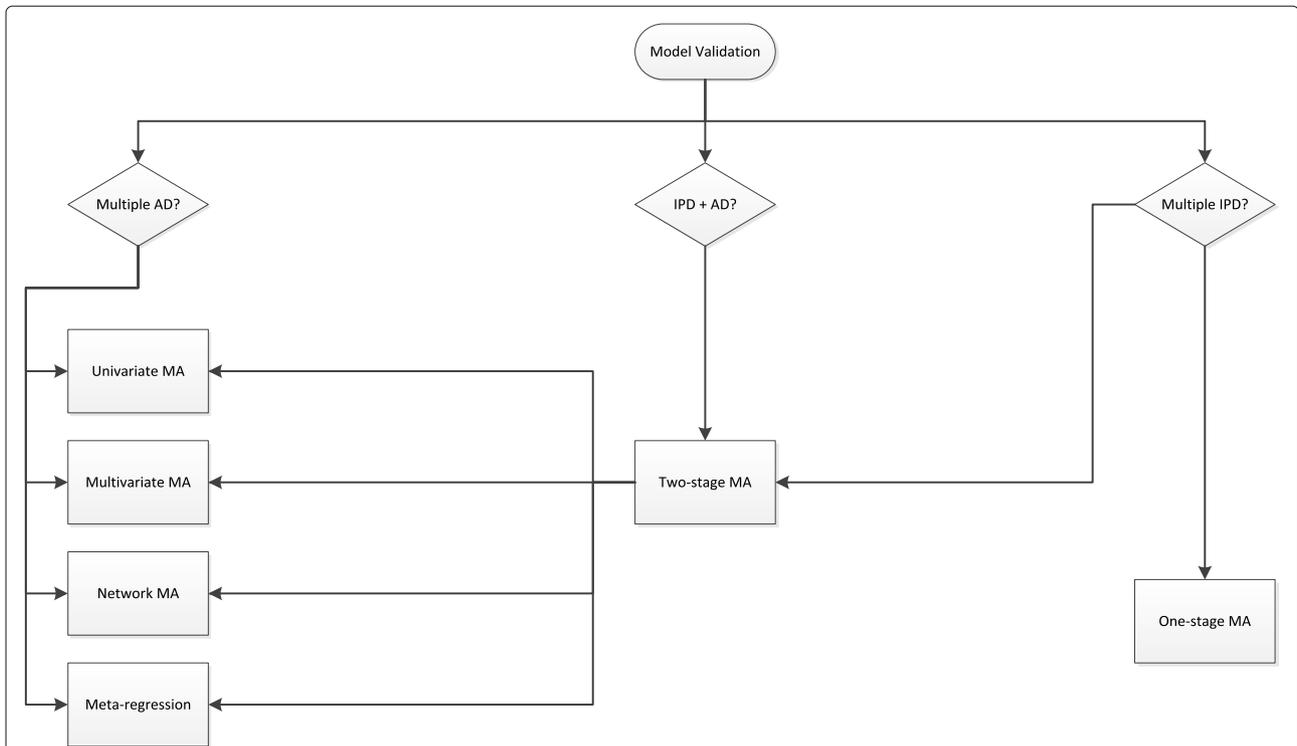
(and their standard errors and within-study correlations), by inconsistent covariate adjustment across studies, and by the presence of between-study heterogeneity. For this reason, meta-analysis of previously developed prognostic models only seems reasonable when the corresponding studies are fairly homogeneous and when the required AD are reported in sufficient detail (see also Fig. 3).

#### **Meta-analysis using IPD**

When IPD are available, it becomes possible to assess and optimize the prognostic model's performance across different settings and populations using a one-stage or a two-stage meta-analysis approach.

#### **Validation of an existing prognostic model**

In the two-stage approach, the model is first validated separately in each IPD, yielding study-specific estimates of model discrimination and calibration. These estimates are then pooled across studies in the second stage, using univariate [66, 70, 72] or multivariate [73] meta-analysis



**Fig. 4** Available methods for quantitative synthesis during prognostic model validation. Abbreviations: MA, meta-analysis; AD, aggregate data; IPD, individual participant data

methods (Fig. 4). For instance, Snell et al. adopted multivariate IPD meta-analysis to summarize the calibration slope and concordance statistic of a prognostic model for breast cancer incidence. The summary estimates were then used in combination with estimates of between-study heterogeneity to calculate the probability that model performance would be adequate (i.e., within certain ranges) in new populations [73].

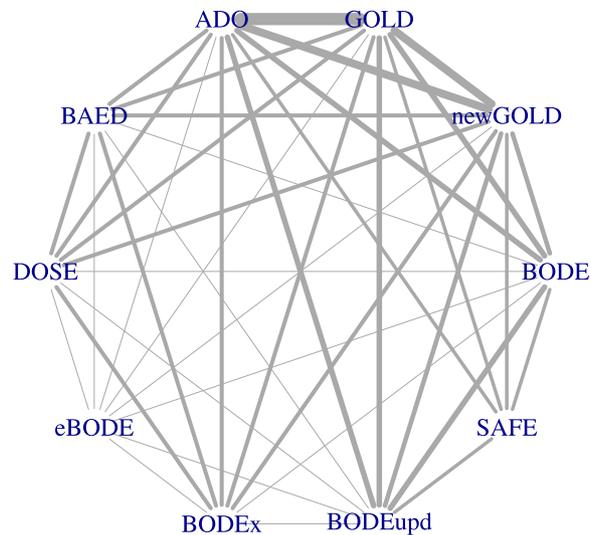
Model validation can also be performed through a one-stage approach. For instance, the summary calibration slope can be derived by fitting a mixed effect model with study-specific intercept terms and a random effect for the prognostic index.

Finally, several extensions of one-stage and two-stage meta-analysis are possible. For instance, network meta-analysis (NMA) can be used to assess the (relative) performance of multiple prognostic models [74], which is particularly helpful when direct comparisons are not feasible for some studies. As an example, Haile et al. compared the performance of 10 prognostic models for calculating mortality risk in patients with chronic obstructive pulmonary disease [74]. Although IPD were available for 24 cohort studies ( $N = 15\,762$ ), information on important variables was often missing such that some models could not be validated in one or more studies (Fig. 5). A two-stage NMA was therefore adopted to summarize all available evidence on the models' comparative

performance and to allow the inclusion of studies where only few models could be validated.

#### **Development of a new prognostic model**

Meta-analysis of IPD is used increasingly often to develop new prognostic models, with improved generalizability across different settings and populations. Meta-analysis approaches are similar to prognostic factor research, and may involve a one-stage or a two-stage approach (see also Fig. 3) [70]. In the two-stage approach, the parameters of the prognostic model (e.g. intercept term and regression coefficients) are estimated separately in each study and subsequently combined across studies using either a fixed or random effects meta-analysis. Conversely, in the one-stage approach, all IPD are simultaneously analyzed by assuming a common, fixed, or random effect for each model parameter. Both approaches then yield a set of study-specific and/or "pooled" regression coefficients that can be used for making absolute risk predictions in a variety of populations. One-stage approaches are particularly helpful when studies are relatively small, or contain few events, as they use a more exact statistical approach and do not require continuity corrections when (partial) separation occurs [38]. Conversely, two-stage approaches are generally preferred when modeling interactions or non-linear terms, as they guard against over-parameterization and ecological bias [43].



**Fig. 5** Validation of 10 prognostic models for 3-year mortality in patients with chronic obstructive pulmonary disease. Depiction of network structure with lines weighted by the total number of participants available for each model comparison [74]. Abbreviations: GOLD, Global initiative for chronic Obstructive Lung Disease; BODE, Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODE upd., BODE updated; ADO, Age, Dyspnoea, airflow Obstruction (we use the updated version of the ADO score in our analysis); e-BODE, severe acute exacerbation of COPD plus BODE; BODEx, Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE, Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE, Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D, Body-mass index, Acute Exacerbations, Dyspnoea

As an example, Westeneng et al. recently performed a meta-analysis with IPD from 14 European cohorts to develop the ESCALC model for predicting survival in patients with amyotrophic lateral sclerosis [75]. They fitted a Royston-Parmer survival model in the entire set of  $N = 11475$  patients and assumed a common baseline hazard and regression coefficients across cohorts. Because the resulting model showed some extent of mis-calibration upon validation, recalibrated cohort-specific baseline hazard functions were reported to enable researchers to tailor model predictions to their population.

A particular advantage of IPD meta-analysis is that it enables the direct evaluation and optimization of a model's generalizability across different settings and populations through internal-external cross-validation [64, 65, 76–78]. Briefly, this method iteratively omits one study from the meta-analysis to externally validate a model that is developed on the remaining studies. This process is repeated several times, leading to multiple estimates of model performance, which in turn can be summarized using aforementioned meta-analysis methods [68, 73]. If performance appears adequate across the available studies, the pooled data is used to develop a final model. Otherwise, it flags heterogeneous study populations where a developed model might not perform well and signals that additional predictors or more advanced modeling approaches (such as the inclusion of

non-linear terms) or updating strategies (such as recalibration) might be needed.

Internal-external cross-validation has, for instance, been adopted during the development of ESCALC, a prognostic model for predicting survival in patients with amyotrophic lateral sclerosis. A one-stage approach was used to estimate a Royston-Parmer model using IPD from all but one study, after which its external validity was evaluated in the omitted study. The process was repeated for all studies, providing 14 estimates of discrimination and calibration performance. These estimates were then pooled using a random effects meta-analysis, yielding a summary c-statistic and calibration slope of, respectively, 0.78 (95% PI 0.74 to 0.82) and 1.01 (95% PI 0.83 to 1.18). These results suggest that the model is likely to perform well across different settings and populations.

#### Meta-analysis using IPD and AD

##### Validation of an existing prognostic model

Because IPD is commonly unavailable for one or more relevant validation studies, researchers may consider a two-stage meta-analysis to combine published estimates of prediction model performance with those derived from the IPD at hand. This approach has, however, not extensively been studied yet, and further research is also warranted to explore alternative strategies such as hierarchical-related regression.

### **Development of a new prognostic model**

For many disease areas, there is an abundance of competing models that predict similar outcomes in related populations. Hence, rather than developing a new prognostic model from scratch, it can be advantageous to combine the AD of the existing models with the available IPD [79–82]. One approach is to summarize the models' regression coefficients together with the associations from the IPD [51, 52]. This is particularly useful if the data are reasonably homogeneous, as synthesis then yields a prognostic model that is applicable to the "average" population. Conversely, when studies have different baseline risk or predictor-outcome associations, some tailoring will often be necessary to ensure that the new model remains sufficiently accurate in local settings. In these situations, the IPD can be used to adjust the existing models to specific populations by adopting Bayesian inference [52], model averaging [81], regression analysis [79, 81, 83, 84], or mixture models [83].

For example, model averaging was recently applied to combine the logistic EuroSCORE and EuroSCORE II models for predicting short-term mortality in patients undergoing coronary artery bypass graft surgery [82]. These models showed substantial mis-calibration in contemporary registry data and were therefore combined into a single model that was tailored to the contemporary population.

### **Summary points**

Many prognostic model studies are based on relatively small samples, leading to overfitting, poor generalizability, and over-optimism [58, 85]. Evidence synthesis allows to increase the effective sample size and to study more diverse settings and populations [62, 64]. Although synthesis is ideally based on IPD, a systematic review and meta-analysis of published data can initially be performed to study the (discrimination and calibration) performance of a previously developed model. Estimates of between-study heterogeneity can then help to reveal the extent of necessary improvements (e.g., local tailoring) and to calculate the probability that the model(s) will be clinically useful in certain settings [73, 86]. In general, a good model will have satisfactory performance across different settings and populations. However, if prediction model performance is poor overall or prone to substantial between-study heterogeneity, retrieval of IPD may help to study causes of detrimental performance [66, 67, 87] and to establish whether distinct models are needed for different settings and populations [61].

When developing new or updating existing models, it is important to consider heterogeneity in baseline risk, predictor effects, the linear predictor, and the absolute risk

predictions [61]. Risk predictions should be reasonably similar across studies for a prediction model to be labeled "generalizable," and therefore, it is helpful to limit any heterogeneity in baseline risk and predictor effects while keeping the model's overall performance sufficiently high. Although internal-external cross-validation using IPD from multiple studies may be helpful to achieve this, further research is needed to integrate this endeavor in a statistical framework.

Finally, for newly developed prediction models from IPD-MA, it is helpful to provide any information that allows for tailored predictions. For instance, appropriate intercept terms can often be derived from the outcome incidence, particularly if predictor variables have been centered around their local means [77]. Similarly, predictor effects can sometimes be tailored using information about their particular measurement [88]. When it remains unclear which parameter values (e.g., intercept term) are most appropriate for predictions in new populations, researchers may use the pooled estimates or, preferably, integrate over the distribution of the random effects [89].

### **Concluding remarks**

In this paper, we have summarized and sign-posted various methods for meta-analysis of prognostic factor and prognostic model studies. Because these primary prognosis studies may address very different types of research questions and are often poorly reported, advanced meta-analysis methods are usually needed to provide (meaningful) summary estimates and understand sources of between-study heterogeneity. Regardless, researchers should not be daunted by their complexity, as we have shown that many of these methods have been implemented in traditional software packages and lead to an improved understanding of prognosis-related research questions.

For researchers embarking on a meta-analysis, the following issues should be taken into account. First, it is important to ensure that available data are of sufficient relevance and quality. It is recommended to conduct a systematic review of the literature and to harmonize available IPD sets. Similarity of datasets can, for instance, be improved by standardizing related measurement scales [90], by adopting measurement error correction methods [91–93], or by treating bias arising from measurement error as a missing data problem [90, 92, 94]. Second, when datasets are affected by missing data, advanced imputation methods are needed to ensure valid inferences [95–97]. Finally, it is important to realize that not all meta-analysis methods have yet been rigorously assessed and that further research is still needed to explore their potential areas of application.

### Abbreviations

AD: Aggregate data; CI: Confidence interval; CIMT: Carotid intima-media thickness; FP: Fractional polynomial(s); HR: Hazard ratio; IPD: Individual participant data; IPD-MA: Individual participant data meta-analysis; NMA: Network meta-analysis; TBI: Traumatic brain injury

### Acknowledgements

We would like to thank two anonymous referees whose helpful comments have greatly improved this article.

### Funding

While undertaking this work, TD was supported by the Netherlands Organisation for Health Research and Development (91617050 and 91215058).

### Availability of data and materials

Not applicable.

### Authors' contributions

TD drafted the initial manuscript. All authors provided feedback on the text, methods, and examples. TD and VDJ revised the manuscript according to the comments from the co-authors. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable.

### Competing interests

TPD is an Associate Editor for BMC Diagnostic and Prognostic Research. KGM is an Editor-in-Chief for BMC Diagnostic and Prognostic Research. RDR is a member of the Editorial Board of BMC Diagnostic and Prognostic Research.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. <sup>2</sup>Cochrane Netherlands, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. <sup>3</sup>Research Institute for Primary Care & Health Sciences Keele University, Staffordshire, ST5 5BG, The UK.

Received: 23 July 2018 Accepted: 16 April 2019

Published online: 11 July 2019

### References

- Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. *BMC Med Res Methodol*. 2009;9(1):29. <https://doi.org/10.1186/1471-2288-9-29>.
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71–2.
- Debray TPA, Moons KGM, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RHH, Reitsma JB. GetReal methods review group: get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods*. 2015;6:239–309. <https://doi.org/10.1002/jrsm.11>.
- Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet*. 1998;351(9096):123–7. [https://doi.org/10.1016/S0140-6736\(97\)08468-7](https://doi.org/10.1016/S0140-6736(97)08468-7).
- Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ*. 1997;315(7119):1371–4. <https://doi.org/10.1136/bmj.315.7119.1371>.
- Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis Research in Healthcare: concepts, methods, and impact. Oxford: Oxford University Press; 2019.
- Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KGM, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD. PROGRESS Group: Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;e5595:346.
- Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H, for the PROGRESS Group: Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380. <https://doi.org/10.1371/journal.pmed.1001380>.
- Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, for the PROGRESS Group: Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. <https://doi.org/10.1371/journal.pmed.1001381>.
- Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KGM, Steyerberg EW, Schroter S, Sauerbrei W, Altman DG, Hemingway H. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ*. 2013;e5793:346. <https://doi.org/10.1136/bmj.e5793>.
- Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet*. 1995;345(8965):1616–9.
- Damen JAAG, Hoof L. The increasing need for systematic reviews of prognosis studies: strategies to facilitate review production and improve quality of primary research. *Diagn Prognostic Res*. 2019;3(1):2. <https://doi.org/10.1186/s41512-019-0049-6>.
- Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ*. 2001;323(7306):224–8.
- Moons KGM, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of clinical prediction modelling studies: The CHARMS Checklist. *PLoS Med*. 2014;11(10):e1001744. <https://doi.org/10.1371/journal.pmed.1001744>.
- Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280–6. <https://doi.org/10.7326/0003-4819-158-4-201302190-00009>.
- Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:375. <https://doi.org/10.1136/bmj.b375>.
- Sauerbrei W, Holländer N, Riley RD, Altman DG. Evidence-based assessment and application of prognostic markers: the long way from single studies to meta-analysis. *Commun Stat Theory Methods*. 2006;35(7):1333–42. <https://doi.org/10.1080/03610920600629666>.
- Riley RD, Moons K, Snell KIE, Ensor J, Hoof L, Altman D, Hayden J, Collins G, Debray T. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ*. 2019;k4597:364. <https://doi.org/10.1136/bmj.k4597>.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97–111. <https://doi.org/10.1002/jrsm.12>.
- van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Stat Med*. 2002;21(4):589–624.
- Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A (Stat Soc)*. 2018;181(1):205–27. <https://doi.org/10.1111/rssa.12275>.
- Langan D, Higgins JPT, Simmonds M. Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies. *Res Synth Methods*. 2017;8(2):181–98. <https://doi.org/10.1002/jrsm.1198>.
- Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JPT, Langan D, Salanti G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55–79. <https://doi.org/10.1002/jrsm.1164>.
- Jackson D, Law M, Rücker G, Schwarzer G. The Hartung-Knapp modification for random-effects meta-analysis: a useful refinement but are there any residual concerns? *Stat Med*. 2017;36(25):3923–34. <https://doi.org/10.1002/sim.7411>.
- Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. 2001;20(24):3875–89.
- Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21(21):3153–9. <https://doi.org/10.1002/sim.1262>.
- Zhang Y, Zhao D, Gong C, Zhang F, He J, Zhang W, Zhao Y, Sun J. Prognostic role of hormone receptors in endometrial cancer: A systematic review and meta-analysis. *World J Surg Oncol*. 2015;13:208. <https://doi.org/10.1186/s12957-015-0619-1>.
- Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, White IR. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*. 2017;j3932:358. <https://doi.org/10.1136/bmj.j3932>.

29. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med*. 2011;30(20):2481–98. <https://doi.org/10.1002/sim.4172>.
30. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982–90. <https://doi.org/10.1016/j.jclinepi.2005.02.022>.
31. Riley RD. Multivariate meta-analysis: The effect of ignoring within-study correlation. *J R Stat Soc Series B Stat Methodol*. 2009;172(4):789–811. <https://doi.org/10.1111/j.1467-985X.2008.00593.x>.
32. Yoneoka D, Henmi M. Synthesis of linear regression coefficients by recovering the within-study covariance matrix from summary statistics. *Res Synth Methods*. 2017;8(2):212–9. <https://doi.org/10.1002/jrsm.1228>.
33. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172–86. <https://doi.org/10.1093/biostatistics/kxm023>.
34. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res Synth Methods*. 2010;1(1):2–19. <https://doi.org/10.1002/jrsm.4>.
35. Riley RD, Elia EG, Malin G, Hemming K, Price MP. Multivariate meta-analysis of prognostic factor studies with multiple cut-points and/or methods of measurement. *Stat Med*. 2015;34(17):2481–96. <https://doi.org/10.1002/sim.6493>.
36. Shi JQ, Copas JB. Meta-analysis for trend estimation. *Stat Med*. 2004;23(1):3–19. <https://doi.org/10.1002/sim.1595>.
37. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology*. 1993;4(3):218–28.
38. Debray TPA, Moons KGM, Abo-Zaid GMA, Koffijberg H, Riley RD. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS ONE*. 2013;8(4):e60650. <https://doi.org/10.1371/journal.pone.0060650>.
39. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med*. 2017;36(5):855–75. <https://doi.org/10.1002/sim.7141>.
40. Trivella M, Pezzella F, Pastorino U, Harris AL, Altman DG. Prognosis In Lung Cancer (PILC) Collaborative Study Group: Microvessel density as a prognostic factor in non-small-cell lung carcinoma: a meta-analysis of individual patient data. *Lancet Oncol*. 2007;8(6):488–99. [https://doi.org/10.1016/S1470-2045\(07\)70145-6](https://doi.org/10.1016/S1470-2045(07)70145-6).
41. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Stat Med*. 2011;30(28):3341–60. <https://doi.org/10.1002/sim.4333>.
42. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med*. 2004;23(16):2509–25. <https://doi.org/10.1002/sim.1815>.
43. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med*. 2012;31(29):3821–39. <https://doi.org/10.1002/sim.5471>.
44. Abo-Zaid G, Sauerbrei W, Riley RD. Individual participant data meta-analysis of prognostic factor studies: state of the art? *BMC Med Res Methodol*. 2012;12:56. <https://doi.org/10.1186/1471-2288-12-56>.
45. Abo-Zaid G, Guo B, Deeks JJ, Debray TPA, Steyerberg EW, Moons KGM, Riley RD. Individual participant data meta-analyses should not ignore clustering. *J Clin Epidemiol*. 2013;66(8):865–873. <https://doi.org/10.1016/j.jclinepi.2012.12.017>.
46. Bowden J, Tierney JF, Simmonds M, Copas AJ, Higgins JP. Individual patient data meta-analysis of time-to-event outcomes: one-stage versus two-stage approaches for estimating the hazard ratio under a random effects model. *Res Synth Methods*. 2011;2(3):150–62. <https://doi.org/10.1002/jrsm.45>.
47. Siannis F, Barrett JK, Farewell VT, Tierney JF. One-stage parametric meta-analysis of time-to-event outcomes. *Stat Med*. 2010;29(29):3030–45. <https://doi.org/10.1002/sim.4086>.
48. Den Ruijter HM, Peters SAE, Anderson TJ, Britton AR, Dekker JM, Eijkemans MJ, Engström G, Evans GW, de Graaf J, Grobbee DE, Hedblad B, Hofman A, Holewijn S, Ikeda A, Kavousi M, Kitagawa K, Kitamura A, Koffijberg H, Lonn EM, Lorenz MW, Mathiesen EB, Nijpels G, Okazaki S, O'Leary DH, Polak JF, Price JF, Robertson C, Rembold CM, Rosvall M, Rundek T, Salonen JT, Sitzer M, Stehouwer CDA, Witteman JC, Moons KG, Bots ML. Common carotid intima-media thickness measurements in cardiovascular risk prediction: A meta-analysis. *JAMA*. 2012;308(8):796–803. <https://doi.org/10.1001/jama.2012.9630>.
49. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Bouillon F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*. 2008;27(11):1870–93. <https://doi.org/10.1002/sim.3165>.
50. Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J Clin Epidemiol*. 2007;60(5):431–9. <https://doi.org/10.1016/j.jclinepi.2006.09.009>.
51. Yoneoka D, Henmi M, Sawada N, Inoue M. Synthesis of clinical prediction models under different sets of covariates with one individual patient data. *BMC Med Res Methodol*. 2015;15(1):101. <https://doi.org/10.1186/s12874-015-0087-x>.
52. 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*. 2013;32(23):2697–712. <https://doi.org/10.1002/sim.5412>.
53. Debray TPA, Koffijberg H, Lu D, Vergouwe Y, Steyerberg EW, Moons KGM. Incorporating published univariable associations in diagnostic and prognostic modeling. *BMC Med Res Methodol*. 2012;12:121. <https://doi.org/10.1186/1471-2288-12-121>.
54. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med*. 2000;19(2):141–60. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000130\)19:2<141::AID-SIM334>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0258(20000130)19:2<141::AID-SIM334>3.0.CO;2-O).
55. Claeys KC, Zasowski EJ, Lagnf AM, Levine DP, Davis SL, Rybak MJ. Novel application of published risk factors for methicillin-resistant *S. aureus* in acute bacterial skin and skin structure infections. *Int J Antimicrob Agents*. 2018;51(1):43–6. <https://doi.org/10.1016/j.ijantimicag.2017.05.015>.
56. The Fibrinogen Studies Collaboration. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Stat Med*. 2009;28(8):1218–37. <https://doi.org/10.1002/sim.3540>.
57. Becker BJ, Wu M-J. The Synthesis of Regression Slopes in Meta-Analysis. *Stat Sci*. 2007;22(3):414–29. <https://doi.org/10.1214/07-STS243>.
58. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2014;68(1):25–34. <https://doi.org/10.1016/j.jclinepi.2014.09.007>.
59. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–8. <https://doi.org/10.1136/heartjnl-2011-301247>.
60. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: Validating a prognostic model. *BMJ*. 2009;b605:338. <https://doi.org/10.1136/bmj.b605>.
61. Steyerberg EW, Nieboer D, Debray TPA, Van Houwelingen JC. Assessment of heterogeneity in meta-analysis of prediction models: An overview and illustration. *Stat Med*. 2019.
62. Wynants L, Kent DM, Timmerman D, Lundquist CM, Van Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. *Diagn Prognostic Res*. 2019;3(1):6. <https://doi.org/10.1186/s41512-019-0046-9>.
63. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, Collins GS. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *BMJ*. 2016;i3140:353. <https://doi.org/10.1136/bmj.i3140>.
64. Debray TPA, Riley R, Rovers M, Reitsma JB, Moons K. on behalf of the Cochrane IPD Meta-analysis Methods group: Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):e1001886. <https://doi.org/10.1371/journal.pmed.1001886>.
65. Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. <https://doi.org/10.1186/1471-2288-14-3>.
66. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KGM. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2019;i6460. <https://doi.org/10.1136/bmj.i6460>.

67. Debray TPA, Damen JAAG, Riley R, Snell KIE, Reitsma JB, Hooft L, Collins G, Moons KG. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res In Press*. 2018. <https://doi.org/10.1177/0962280218785504>.
68. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-statistic and calibration measures? *Stat Meth Med Res*. 2017;27(11):3505–22. <https://doi.org/10.1177/0962280217705678>.
69. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29(29):3046–67. <https://doi.org/10.1002/sim.4040>.
70. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Emerging Risk Factors Collaboration: assessing Risk Prediction Models Using Individual Participant Data From Multiple Studies. *Am J Epidemiol*. 2013;179(5):621–32. <https://doi.org/10.1093/aje/kwt298>.
71. van Doorn S, Debray TPA, Kaasenbrood F, Hoos AW, Rutten FH, Moons KGM, Geersing G-J. Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis. 2017;15:1–13. <https://doi.org/10.1111/jth.13690>.
72. van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y. Assessing discriminative ability of risk models in clustered data. *BMC Med Res Methodol*. 2014;14:5. <https://doi.org/10.1186/1471-2288-14-5>.
73. Snell KIE, Hua H, Debray TPA, Ensor J, Look MP, Moons KGM, Riley RD. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2015;69:40–50. <https://doi.org/10.1016/j.jclinepi.2015.05.009>.
74. Haile SR, Guerra B, Soriano JB, Puhan MA. 3CIA collaboration: Multiple score comparison: a network meta-analysis approach to comparison and external validation of prognostic scores. *BMC Med Res Methodol*. 2017;17(1):172. <https://doi.org/10.1186/s12874-017-0433-2>.
75. Westeneng H-J, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, Martin S, McDermott CJ, Thompson AG, Pinto S, Kobeleva X, Rosenbohm A, Stubendorff B, Sommer H, Middelkoop BM, Dekker AM, van Vugt JFA, van Rheenen W, Vajda A, Heverin M, Kazoka M, Hollinger H, Gromicho M, Körner S, Ringer TM, Rödiger A, Gunkel A, Shaw CE, Bredenoord AL, van Es MA, Corcia P, Couratier P, Weber M, Grosskreutz J, Ludolph AC, Petri S, de Carvalho M, Van Damme P, Talbot K, Turner MR, Shaw PJ, Al-Chalabi A, Chiò A, Hardiman O, Moons KGM, Veldink JH, van den Berg LH. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol*. 2018;17(5):423–33. [https://doi.org/10.1016/S1474-4422\(18\)30089-9](https://doi.org/10.1016/S1474-4422(18)30089-9).
76. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2015;69:245–7. <https://doi.org/10.1016/j.jclinepi.2015.04.005>.
77. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med*. 2013;32(18):3158–80. <https://doi.org/10.1002/sim.5732>.
78. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Stat Med*. 2004;23(6):907–26. <https://doi.org/10.1002/sim.1691>.
79. Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of updating clinical prediction models. *Stat Med*. 2017;37:1343–58. <https://doi.org/10.1002/sim.7586>.
80. Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. Clinical prediction in defined populations: a simulation study investigating when and how to aggregate existing models. *BMC Med Res Methodol*. 2017;17(1):1. <https://doi.org/10.1186/s12874-016-0277-1>.
81. Debray TPA, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KGM. Meta-analysis and aggregation of multiple published prediction models. *Stat Med*. 2014;33(14):2341–62. <https://doi.org/10.1002/sim.6080>.
82. Su T.-L., Jaki T, Hickey G, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. *Stat Methods Med Res*. 2018;27(1):185–97. <https://doi.org/10.1177/0962280215626466>.
83. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401–15.
84. Merz CJ, Pazzani MJ. A principal components approach to combining regression estimates. *Mach Learn*. 1999;36:9–32.
85. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, Voysey M, Wharton R, Yu L-M, Moons KG, Altman DG. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40. <https://doi.org/10.1186/1471-2288-14-40>.
86. Wynants L, Riley RD, Timmerman D, Van Calster B. Random-effects meta-analysis of the clinical utility of tests and prediction models. *Stat Med*. 2018;37(12):2034–52. <https://doi.org/10.1002/sim.7653>.
87. Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. *J Clin Epidemiol*. 2016;79:76–85. <https://doi.org/10.1016/j.jclinepi.2016.05.007>.
88. Whittle R, Peat G, Belcher J, Collins GS, Riley RD. Measurement error and timing of predictor values for multivariable risk prediction models are poorly reported. *J Clin Epidemiol*. 2018;102:38–49. <https://doi.org/10.1016/j.jclinepi.2018.05.008>.
89. Pavlou M, Ambler G, Seaman S, Omar RZ. A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes. *BMC Med Res Methodol*. 2015;15:59. <https://doi.org/10.1186/s12874-015-0046-6>.
90. Griffith LE, van den Heuvel E, Fortier I, Soheli N, Hofer SM, Payette H, Wolfson C, Belleville S, Kenny M, Doiron D, Raina P. Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported. *J Clin Epidemiol*. 2015;68(2):154–62. <https://doi.org/10.1016/j.jclinepi.2014.09.003>.
91. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med*. 2014;33(12):2137–55. <https://doi.org/10.1002/sim.6095>.
92. Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. *Stat Methods Med Res*. 2018;27(6):1695–708. <https://doi.org/10.1177/0962280216667764>.
93. Hossain S, Gustafson P. Bayesian adjustment for covariate measurement errors: a flexible parametric approach. *Stat Med*. 2009;28(11):1580–600. <https://doi.org/10.1002/sim.3552>.
94. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol*. 2006;35(4):1074–81. <https://doi.org/10.1093/ije/dyl097>.
95. Audigier V, White IR, Jolani S, Debray TPA, Quartagno M, Carpenter JR, van Buuren S, Resche-Rigon M. Multiple imputation for multilevel data with continuous and binary variables. *Stat Sci*. 2018;33(2):160–83. <https://doi.org/10.1214/18-STS646>.
96. Grund S, Lütke O, Robitzsch A. Multiple imputation of missing data for multilevel models: simulations and recommendations. *Organ Res Methods*. 2018;21(1):111–49. <https://doi.org/10.1177/1094428117703686>.
97. Kunkel D, Kaizar EE. A comparison of existing methods for multiple imputation in individual participant data meta-analysis. *Stat Med*. 2017;36(22):3507–32. <https://doi.org/10.1002/sim.7388>.
98. Kovačič J, Varnai VM. A graphical model approach to systematically missing data in meta-analysis of observational studies. *Stat Med*. 2016;35(4):4443–58. <https://doi.org/10.1002/sim.7010>.