Get real in individual participant data (IPD) meta-analysis: a review of the methodology

Thomas P. A. Debray, a,b* Karel G. M. Moons, a,b
Gert van Valkenhoef, c Orestis Efthimiou, d Noemi Hummel, e
Rolf H. H. Groenwold, a Johannes B. Reitsma a,b and
on behalf of the GetReal methods review group

Individual participant data (IPD) meta-analysis is an increasingly used approach for synthesizing and investigating treatment effect estimates. Over the past few years, numerous methods for conducting an IPD meta-analysis (IPD-MA) have been proposed, often making different assumptions and modeling choices while addressing a similar research question. We conducted a literature review to provide an overview of methods for performing an IPD-MA using evidence from clinical trials or non-randomized studies when investigating treatment efficacy. With this review, we aim to assist researchers in choosing the appropriate methods and provide recommendations on their implementation when planning and conducting an IPD-MA. © 2015 The Authors. Research Synthesis Methods published by John Wiley & Sons, Ltd.

Keywords: meta-analysis; IPD; evidence synthesis; review; RCT; non-randomized intervention studies; NRSI; cross-design

1. Introduction

The evaluation of a novel drug or intervention typically involves a series of randomized clinical trials (RCTs) where its safety and efficacy are extensively tested. Because trials are often relatively small and typically exhibit differences in study design, selection of subjects, studied outcome(s), dosage, choice of comparator intervention, and quality of the conducted research, conflicting evidence occasionally arises. As a consequence, systematic reviews have become an important tool to summarize the evidence from these trials and to generalize their conclusions beyond their specific settings.

Over the past few decades, several methods have been developed to quantify the results of a systematic review. Most of these methods adopt a meta-analytical rationale and pool the results from individual studies by accounting for various forms of uncertainty (Sutton et al., 2009; Sutton and Higgins, 2008). Hereby, group-level summary statistics (aggregate data (AD)) that quantify the treatments’ relative efficacy or safety are retrieved from the published literature or from study authors and are subsequently synthesized into a weighted average.

Keywords: meta-analysis; IPD; evidence synthesis; review; RCT; non-randomized intervention studies; NRSI; cross-design

1. Introduction

The evaluation of a novel drug or intervention typically involves a series of randomized clinical trials (RCTs) where its safety and efficacy are extensively tested. Because trials are often relatively small and typically exhibit differences in study design, selection of subjects, studied outcome(s), dosage, choice of comparator intervention, and quality of the conducted research, conflicting evidence occasionally arises. As a consequence, systematic reviews have become an important tool to summarize the evidence from these trials and to generalize their conclusions beyond their specific settings.

Over the past few decades, several methods have been developed to quantify the results of a systematic review. Most of these methods adopt a meta-analytical rationale and pool the results from individual studies by accounting for various forms of uncertainty (Sutton et al., 2009; Sutton and Higgins, 2008). Hereby, group-level summary statistics (aggregate data (AD)) that quantify the treatments’ relative efficacy or safety are retrieved from the published literature or from study authors and are subsequently synthesized into a weighted average.

Keywords: meta-analysis; IPD; evidence synthesis; review; RCT; non-randomized intervention studies; NRSI; cross-design

1. Introduction

The evaluation of a novel drug or intervention typically involves a series of randomized clinical trials (RCTs) where its safety and efficacy are extensively tested. Because trials are often relatively small and typically exhibit differences in study design, selection of subjects, studied outcome(s), dosage, choice of comparator intervention, and quality of the conducted research, conflicting evidence occasionally arises. As a consequence, systematic reviews have become an important tool to summarize the evidence from these trials and to generalize their conclusions beyond their specific settings.

Over the past few decades, several methods have been developed to quantify the results of a systematic review. Most of these methods adopt a meta-analytical rationale and pool the results from individual studies by accounting for various forms of uncertainty (Sutton et al., 2009; Sutton and Higgins, 2008). Hereby, group-level summary statistics (aggregate data (AD)) that quantify the treatments’ relative efficacy or safety are retrieved from the published literature or from study authors and are subsequently synthesized into a weighted average.
Unfortunately, when synthesizing published AD, even rigorously conducted meta-analyses can be of limited value. In particular, when there is substantial heterogeneity in estimates of relative treatment effect, a weighted average may no longer be informative in medical care. In such situations, it is important to identify whether treatment effects vary across clinical subgroups because of effect modification. Although published AD can be used for exploring modifiers of treatment effect, such approach lacks power when published summary statistics (e.g., mean age) do not vary much across studies. More importantly, the use of published AD to investigate effect modification is prone to bias because it cannot properly take subject-level characteristics into account. This bias is also known as the “ecological fallacy” (Berlin et al., 2002). Additional problems arise when AD are not available, poorly reported, derived and presented differently across studies (for example, odds ratio versus relative risk), and more likely to be reported (and in greater detail) when statistically or clinically significant (Riley et al., 2010). For this reason, investigators increasingly embark into an individual participant data meta-analysis (IPD-MA) (Riley et al., 2010; Thompson, 2009; Stewart and Tierney, 2002). These meta-analyses include the raw data from each relevant study that is (ideally) identified through a systematic review. By securing IPD of individual trials, it becomes possible to disentangle subject-level and study-level sources of heterogeneity in treatment effect (van Walraven, 2010; Michiels et al., 2005; Lyman and Kuderer, 2005). This, in turn, may help to explore effect modification (Simmonds and Higgins, 2007; Stewart and Tierney, 2002) and to consistently adjust for confounding variables (i.e., for differences in baseline characteristics across different treatment arms). Access to IPD may also help to improve data quality (Lyman and Kuderer, 2005; Stewart and Tierney, 2002; Lau et al., 1998), to standardize definitions and analyses (van Walraven, 2010; Lyman and Kuderer, 2005), to obtain complete follow-up data on all randomized participants (Clarke and Stewart, 1994), to combine studies with different follow-up times (Sud and Douketis, 2009; Lyman and Kuderer, 2005; Stewart and Tierney, 2002; Lau et al., 1998), to analyze multiple outcomes (Bujkiewicz et al., 2014; Li and Meredith, 2003), to investigate long-term outcomes (Stewart and Tierney, 2002), and to investigate rare exposures.

Because limited guidance currently exists on dealing with challenges that are specific to IPD-MA, we embarked on a literature review to identify state-of-the-art methods and to provide recommendations on their implementation. This review aims to help systematic reviewers with a limited background in medical statistics in identifying relevant methods for their IPD-MA.

2. Methods

2.1. Information sources

We conducted a literature review to identify scientific, peer-reviewed articles that focus on methods for IPD-MA to investigate treatment efficacy using evidence from clinical trials or non-randomized studies. Medline and Embase databases were queried using a combination of relevant keywords in the Pubmed and Ovid platform (Supporting Information 2). The corresponding search strategy was independently reviewed by two librarians. They concluded that the search strategy did not meaningfully improve when medical subject headings terms were included or search terms were truncated. A selected set of key journals was also hand-searched: the Journal of Research Synthesis Methods and the Journal of the Royal Statistical Society (series A, B, and C). Databases were searched from the earliest date until January 27, 2014. Finally, a last attempt to identify missed key articles was made by contacting key researchers in the field and performing cross-reference checks. Here, relevant hits were added until March 1, 2015.

Because the search for IPD methods was limited to medical and natural science databases, this review is not intended to be comprehensive. It is, for instance, well known that several IPD-MA methods have been developed in the field of social sciences. In general, the concepts of meta-analysis are universally applicable across different scientific domains, whereas the nature of the data may require tailoring of statistical methods. As a consequence, a more extensive search strategy is unlikely to add relevant variations in the identified methods and may fail to yield any new insights. This effect is known as theoretical saturation (Lilford et al., 2001).

2.2. Eligibility criteria

Articles were considered eligible for inclusion if they were written in English and addressed issues related to IPD-MA in intervention research. Key IPD-MA issues were defined as statistical models, software routines, simulation studies, empirical comparisons, didactic examples, and best practice guidelines. Articles were excluded when IPD-MA methods were applied to address solely a substantive clinical question, without a methodological focus. Articles for which no full text was available were also excluded.

2.3. Search results

A flowchart of the articles identified from our search is presented in Figure 1. A total of 3360 unique records were identified using the search strategy and were deemed eligible for title screening. Out of these, 3206 records were excluded because they did not have a methodological focus or did not correspond to a peer-reviewed
article (e.g., conference poster). From the remaining 154 articles, another 16 records were excluded because of missing full text or non-English language. Finally, when assessing full text, 16 records were added after cross-reference checks, and one article was removed because it did not adhere to inclusion criteria. A total of 153 articles were included in the review (Supporting Information 4); a complete list of the included articles can be accessed online at https://www.zotero.org/groups/wp4_-_ipd_meta-analysis.

3. Results

3.1. Conceptual issues

The IPD from a single RCT are often analyzed using regression models where the observed outcome is modeled as a function of the subjects’ undergone treatment (and, in many cases, other covariates of interest). The intercept term of these models then represents the study effect (e.g., baseline risk), whereas their regression coefficient represents the treatment effect. Regression models have two major advantages: First, their implementation is possible for numerous outcome types (e.g., continuous, binary, and time-to-event), and second, their estimated coefficients can be interpreted in a fairly straightforward manner. Unfortunately, the implementation of traditional regression models becomes challenging when a trial consists of multiple centers or when multiple trials need to be analyzed (as in an IPD-MA). In those situations, regression models need to account for clustering of subjects within trials (or within centers) by allowing each trial (or center) to have their “own” study and treatment effect. This approach becomes infeasible when individual trials (or centers) are relatively small. Moreover, corresponding regression models no longer yield an overall estimate of relative treatment effect, casting doubt about whether and under what circumstances a certain intervention is advantageous.
Meta-analysis attempts to overcome these issues by pooling relative treatment effects from multiple trials or centers (henceforth studies) and by borrowing information across these studies during the process. Over the past few decades, two alternate approaches have been proposed for conducting an IPD-MA (Stewart et al., 2012; Riley et al., 2010). The most common and conceptually least complicated approach is the so-called two-stage approach. In this approach, the IPD are first analyzed separately in each study to produce study-specific estimates of relative treatment effect (and possible treatment-covariate interactions). A combined estimate is then obtained in the second step by calculating a weighted average (e.g., inverse error-variance-based) of the individual estimates. Hereby, one may assume that all studies share a common treatment effect and that differences between estimates from the included studies solely arise because of sampling variation (fixed effects meta-analysis). Alternatively, it is possible to allow for heterogeneous treatment effects by incorporating variation between studies and performing a random effects meta-analysis.

In the so-called one-stage approach, the IPD from all studies are analyzed simultaneously by adopting a single statistical model that fully accounts (random effects meta-analysis), partially accounts (mixed effects meta-analysis), or does not account (fixed effects meta-analysis) for heterogeneity across studies. These single-stage models are more commonly known as multilevel or hierarchical models (Simmonds et al., 2005). The implementation of some common one-stage and two-stage meta-analysis model is illustrated in Supporting Information 3.

It is conventionally believed that the one-stage and two-stage approaches yield similar estimates of treatment effects (Stewart et al., 2012; Koopman et al., 2008b; Tudur Smith and Williamson, 2007; Steinberg et al., 1997). Indeed, it has been shown that under certain conditions, the one-stage and two-stage approaches may lead to equivalent results (Lyman and Kuderer, 2005; Mathew and Nordström, 2010), particularly when interest lies in estimating a single treatment effect estimate. The two-stage approach offers several advantages favoring its implementation in the medical literature. In particular, the two-stage approach is conceptually more intuitive and therefore requires less statistical expertise when conducting or interpreting an IPD-MA (Stewart et al., 2012; Koopman et al., 2007). Furthermore, because the two-stage approach analyzes each study separately, study-specific estimates are not influenced by external information. This discrepancy may, for instance, become relevant when effect modification varies across studies and corresponding associations do not follow a well-defined distribution (Piepho et al., 2012).

Unfortunately, the two-stage approach is known to have little power for detecting nonlinear associations between continuous exposures and the outcome(s) of interest and for detecting treatment-covariate interactions (Simmonds and Higgins, 2007; Simmonds et al., 2005; Tudur Smith et al., 2005b; Schmid et al., 2004). In addition, the two-stage approach may lead to bias in pooled effects, standard errors, between-study heterogeneity, and correlation between random effects when few studies or few participants (or events) per study are available (Debray et al., 2013; Stijnen et al., 2010; Tobias et al., 2004), when statistical models cannot fully account for follow-up times (Poppe et al., 2011; Lyman and Kuderer, 2005; Duchateau et al., 2001; Buyse and Piedbois, 1996) or for the time between recurrent events (Haines and Hill, 2011; Jones et al., 2009). Although methodology to overcome some of these limitations has been described (Stijnen et al., 2010), the one-stage approach offers the highest degree of flexibility for making necessary assumptions and is therefore often considered to be superior. The two-stage approach can, however, still be useful to explore the available data, to present intermediate results or to identify key challenges when designing a one-stage meta-analytical model (Stewart et al., 2012).

Although IPD-MA is often considered as gold standard, their implementation is no panacea against the limitations of meta-analyses that are solely based on published AD. In particular, meta-analysis of raw data may still be difficult when randomization or follow-up procedures vary, when studies have measured different covariates or when relevant descriptions of samples, settings, and treatments are not available for every study included in the meta-analysis (Burdett and Stewart, 2002). Furthermore, IPD-MA may become prone to bias when IPD are only sought for a specific subset of studies (e.g., studies conducted by contacts or friends in the research field) or when the availability of IPD is related to study results (Ahmed et al., 2012).

Several recommendations have been made to accommodate for these potential sources of bias (Ahmed et al., 2012; Stewart and Tierney, 2002). First of all, meta-analyses should be based on studies that are identified through a systematic review. Secondly, clinicians from the field should actively be involved as they may help to identify whether important studies may have been missed (Clarke and Godwin, 1998), to collect additional follow-up data where relevant and to resolve inconsistencies between different data sources. Finally, researchers should seek IPD for all relevant studies identified and collect AD for studies where IPD could not be obtained. Methods for combining IPD and AD are discussed in Section 3.5.

### 3.2. Statistical models to estimate an overall summary of treatment effect

In the succeeding texts, we describe the common methodology of one-stage IPD-MA models. These hierarchical models can be viewed as a direct extension of traditional regression models and are also known as generalized linear mixed models (GLMMs). In contrast to traditional regression models, hierarchical regression models allow coefficients to follow a certain distribution. For instance, GLMMs may specify that the relative treatment effect between two specific interventions varies across studies according to a normal distribution. The mean of this
distribution then simply represents the “average” treatment effect between the two interventions, and its variance indicates the degree of between-study heterogeneity in treatment effect.

Consider an IPD-MA of \( i = 1, \ldots, M \) independent studies with \( k = 1, \ldots, N_i \) subjects each. Let \( x_{ik} \) be a dummy variable that indicates treatment group (treatment or control) of subject \( k \) in study \( i \). We denote the corresponding outcome as \( y_{ik} \). This outcome can be continuous, binary, ordinal, count, and so on. The GLMM can then be stated as follows:

\[
g(E(y_{ik})) = \alpha + \beta_i x_{ik} \tag{1}
\]

In this model, \( E(y_{ik}) \) denotes the expected value of \( y_{ik} \), the parameter \( \alpha \) represents the study effect (e.g., baseline risk), and the parameter \( \beta_i \) represents the treatment effect. The link function \( g \) may take different forms depending on the type of outcome data. A detailed overview of possible implementations is discussed in Section 3.4 and also illustrated in Table 1.

When specifying the study effects in model (1), each \( \alpha_i \) may be taken as fixed effects (estimated separately in each study), as a common effect (so \( \alpha_i = \alpha \) for all studies) or as random effects (\( \alpha_i \) is drawn from a certain distribution). Researchers typically allow for heterogeneous study effects by estimating a separate intercept \( \alpha_i \) in each study (Abo-Zaid et al., 2013; Higgins et al., 2001; Thompson et al., 2000; Turner et al., 2000) or by specifying a random-effects distribution on \( \alpha \) (Broström and Holmberg, 2011; Simmonds and Higgins, 2007; Goldstein et al., 2002; Thompson et al., 2001; Turner et al., 2000). A random-effects distribution may also be specified for \( \beta_i \) when heterogeneity in treatment effects across studies is plausible (Abo-Zaid et al., 2013; Bowden et al., 2011; Katsahian et al., 2008; Goldstein et al., 2002; Higgins et al., 2001; Thompson et al., 2001). It is conventional and computationally efficient to assume normal distributions for the random effects (Sutton and Higgins, 2008; Thompson et al., 2001). Other distributions are also possible (Broström and Holmberg, 2011; Thompson et al., 2001) but often require more computational power and occasionally lead to convergence issues.

### Table 1. Basic statistical models for estimating overall treatment effect

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Model type</th>
<th>Basic statistical model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>GLMM</td>
<td>( y_{ik} \sim N(\mu_{ik}, \sigma^2) ) or ( N(\mu_{ik}, \sigma^2) ) (L1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_{ik} = \alpha + \beta_i x_{ik} )</td>
</tr>
<tr>
<td>Binary</td>
<td>GLMM</td>
<td>( y_{ik} \sim \text{Bernoulli}(p_{ik}) ) (L2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \logit(p_{ik}) = \alpha + \beta_i x_{ik} )</td>
</tr>
<tr>
<td>Ordinal</td>
<td>GLMM</td>
<td>( y_{ijk} \sim \text{Bernoulli}(q_{ijk}) ) (L3a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \logit(q_{ijk}) = \alpha + \beta_i x_{ik} )</td>
</tr>
<tr>
<td>Count</td>
<td>GLMM</td>
<td>( y_{ik} \sim \text{Poisson}(\mu_{ik}) ) (L4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \ln(\mu_{ik}) = \alpha + \beta_i x_{ik} )</td>
</tr>
<tr>
<td>Time-to-event</td>
<td>Cox PH</td>
<td>( h_{ik}(t) = h_{0}(t) \exp(\beta_i x_{ik}) ) (L5a)</td>
</tr>
<tr>
<td></td>
<td>Cox PH</td>
<td>( h_{ik}(t) = h_{0}(t) \exp(\alpha_i + \beta_i x_{ik}) ) (L5b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>where ( \alpha_i \sim N(0, \sigma^2) ) or ( \exp(\alpha_i) \sim \sigma^2 \Gamma(\sigma^{-2}) )</td>
</tr>
<tr>
<td></td>
<td>Cox PH</td>
<td>( h_{ik}(t) = h_{0}(t) \exp(\alpha_i + \beta_i x_{ik}) ) (L5c)</td>
</tr>
<tr>
<td></td>
<td>GLMM</td>
<td>( y_{ijk} \sim \text{Poisson}(\mu_{ijk}) ) (L5d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \ln(\mu_{ijk}) = \alpha_i + \beta_i x_{ik} + \lambda_j + \ln(t_{ijk}) )</td>
</tr>
</tbody>
</table>

GLMM, generalized linear mixed model; PH, proportional hazards.

Overview of statistical models for different outcome types. Each model is discussed in Section 3.4.
For instance, consider an IPD-MA with a binary outcome where study effects and treatment effects differ across studies. The following GLMM could then be specified to estimate an overall treatment effect that is adjusted for the presence of between-study heterogeneity:

\[ y_{ik} \sim \text{Bernoulli}(p_{ik}) \]
\[ \logit(p_{ik}) = a_i + \beta_i x_{ik}. \]
\[ \beta_i \sim N(\mu, \tau) \]

In this model, the pooled treatment effect (i.e., the “average” treatment effect across studies) is given by \( \mu \), and the degree of between-study heterogeneity in treatment effect is given by \( \tau \). Because the study effects are estimated separately in each study, there is no pooled study effect.

### 3.3. Statistical models to investigate heterogeneity in treatment effect across and within studies

In the previous section, we described common methods to produce an overall summary of treatment effect. These methods make little or no attempt to explain or investigate differences between study results (heterogeneity) across studies or to identify subgroup effects (Fisher et al., 2011; Sutton and Higgins, 2008; Simmonds and Higgins, 2007). There has been increasing interest in using meta-analysis to go beyond estimating only the overall treatment effect. As a consequence, meta-analyses increasingly attempt to identify modifiers of treatment effect by exploring the presence of interaction effects.

Traditional approaches such as meta-regression and subgroup analysis investigate the presence of trial-level interaction, that is, interaction between treatment status and a specific study-level covariate. This covariate may represent a certain study characteristic (such as level of blinding) or a summarized subject-level characteristic (such as mean age) (Fisher et al., 2011; Sutton et al., 2008; Tudor Smith et al., 2005b; Li and Meredith, 2003; Yamaguchi et al., 2002). For instance, Berlin et al. (2002) used meta-regression to identify whether anti-lymphocyte antibody induction therapy is more beneficial in patients with elevated panel reactive antibodies (PRA). Hereto, they determined the percent of patients with elevated PRA within each study and estimated their association with the corresponding treatment’s effect size (Berlin et al., 2002).

Although trial-level interactions may indicate the presence of effect modification, it has been demonstrated that such interactions have low statistical power for identifying modifiers of treatment effect and may lead to ecological (aggregation) bias. In particular, associations between aggregated values may not be representative for individual subjects (Lambert et al., 2002). In the example of induction, meta-regression failed to identify effect modification by PRA because the level of benefit within PRA category changed across studies. A remedy against this pitfall is to use IPD and to investigate the presence of subject-level interaction (rather than trial-level interaction). This can be achieved by specifying an interaction term between treatment status and subject-level covariate in model (1). Details on how to do this in one-stage and two-stage approaches are provided in Supporting Information 3. The inclusion of interaction terms may also help to adjust for baseline imbalances despite randomization (Higgins et al., 2001) and to increase the precision of treatment effect estimates (by including a strong prognostic factor). Estimated interaction terms must, however, be interpreted with care as they describe a mixture of trial-level and subject-level interactions (Higgins et al., 2001). For this reason, when a modifier of interest has sufficient variation within and across studies, researchers sometimes specify two (instead of one) interaction terms: one interaction between treatment status and the study-specific mean of the covariate (i.e., the values that would be used in meta-regression) and one interaction between treatment status and the study-centered covariate values (Donegan et al., 2012; Fisher et al., 2011). In this manner, it becomes possible to quantify the presence of ecological bias.

Individual participant data meta-analysis often assume common effects for interaction terms to restrict model complexity and facilitate estimation (Crowther et al., 2012; Teramukai et al., 2004). It is, however, possible that interaction terms are actually heterogeneous across studies. For this reason, some researchers propose to estimate subject-level interactions in isolation (“meta-analysis of interaction estimates”). This can be achieved by adopting a two-stage approach where interaction effects are estimated separately within each study and then pooled in the second stage of the IPD-MA (Fisher et al., 2011; Simmonds and Higgins, 2007). Alternatively, it is possible to adopt a one-stage approach by assuming random-effects distributions for the interaction effects (Simmonds and Higgins, 2007; Schmid et al., 2004). The latter strategy has been demonstrated to have the greatest power and flexibility, even when few studies are available, and is therefore generally recommended (Katsahian et al., 2008; Simmonds and Higgins, 2007). Finally, researchers may also attempt to model nonlinear interactions, for example, by using fractional polynomials (Royston and Sauerbrei, 2004), to model two-way interactions between treatment and study-level covariates or even to model three-way interactions between treatment, subject-level covariates, and study-level covariates.

It is important to realize that practical implementation of IPD-MA with interaction terms requires careful thought, as the range of possible analyses (i.e., options for treating the unknown parameters) could easily lead to over-fitting and data-dredging. This problem becomes even more apparent when multiple study-level and...
subject-level covariates are available and IPD are limited. For this reason, it is recommended to pre-specify statistical models in a study protocol before collecting or analyzing the IPD (Higgins et al., 2001).

### 3.4. Modeling of specific types of outcomes

In the succeeding texts, we describe common statistical models for analyzing specific types of outcomes and describe typical estimation procedures. An overview is provided in Table 1.

#### 3.4.1. Continuous.
Continuous outcome data are typically modeled using an identity link function (L1). In this model, \( \beta_i \) indicates the absolute change in outcome due to treatment for patients from the \( i \)th study. If the outcome of interest represents a change from baseline (e.g., a change in depression score since inclusion by the trial), it is appropriate to adjust for baseline responses. Further discussion can be found in Supporting Information 3 and (Riley et al., 2013; Goldstein et al., 2002; Higgins et al., 2001; Thompson et al., 2001; Goldstein et al., 2000).

#### 3.4.2. Binary.
When modeling a binary outcome, it is common to use a logit link function (L2). In this model, \( \exp(\beta_i) \) represents the odds ratio for the treatment effect in the \( i \)th study. If \( \beta_i \) is assumed as a common effect (\( \beta_i = \beta \) for all studies) or as a random effect (e.g., \( \beta_i \sim N(\beta, \tau^2) \)), then \( \exp(\beta) \) represents the odds ratio for the overall treatment effect. Examples are provided by (Thomas et al., 2014; Debray et al., 2013; Stijnen et al., 2010; Sutton et al., 2008; Thompson et al., 2001; Wakefield and Salway, 2001; Turner et al., 2000).

#### 3.4.3. Ordinal.
The logistic model can easily be extended to account for ordinal outcome-type data by relating to the proportional odds model (Thompson et al., 2001; Whitehead et al., 2001). Let \( q_{jk} \) denote the probability of subject \( k \) in study \( i \) having a response in category \( j \) or below. Hereby, it is assumed that the categories are ordered in terms of desirability, thus, lower categories are better. The proportional odds model is described in (L3a), where the parameter \( \xi_j \) represents the log odds ratios for each category cutoff. This model assumes that there is a common treatment effect \( \beta_i \) and a common study effect \( \alpha_i \) within each treatment group. It is possible to relax these assumptions by allowing study effects to vary across different treatment groups (L3b). When the assumption of proportional odds is violated, generalized ordered models or partial proportional models may be used instead of the proportional odds model.

#### 3.4.4. Count.
Denote \( y_{ik} \) as the number of events for subject \( i \) in study \( k \). A straightforward one-stage approach may adopt a Poisson distribution with a log-link (L4). Several extensions have been described that for instance, allow for data with many zeros by introducing a discrete point mass (hierarchical zero-inflated Poisson regression) (Lee et al., 2006; Yau and Lee, 2001; Hall, 2000).

#### 3.4.5. Time-to-event.
Let \( t_{ik} \) denote the observed time (either censoring time or event time) for subject \( k \) in study \( i \). The outcome \( y_{ik} \) then represents an indicator that the time corresponds to an event (\( y_{ik} = 1 \)) or a censoring time (\( y_{ik} = 0 \)). Common multilevel survival models distinguish between heterogeneity in baseline hazard (e.g., due to differences in incidence of the event between studies given the same treatment) and heterogeneity in relative treatment effect (e.g., due to differences in treatment effect size) (Michiels et al., 2005). The hierarchical Cox proportional hazards regression model is a popular approach for analyzing survival data from multiple studies as it requires no assumptions regarding the distribution of the baseline hazard rate (Michiels et al., 2005; Tudor Smith et al., 2005b, 2005a; Sargent, 1998). This model may account for heterogeneity in baseline hazard (that is, the presence of study effects) by stratifying the baseline hazard by study (L5a). Examples of this second strategy are provided in (Crowther et al., 2012; Bowden et al., 2011; Thompson et al., 2010; Katsahian et al., 2008; Tudor Smith and Williamson, 2007; Michiels et al., 2005; Tudor Smith et al., 2005a, 2005b; Trikalinos and Ioannidis, 2001). Because L5a does not yield a direct estimate of the trial effect, more advanced approaches attempt to investigate this heterogeneity by introducing a frailty term \( \zeta \). It is then assumed that the hazards within each study are proportional to the same common baseline hazard function (L5b). Examples of this second strategy are provided in (Crowther et al., 2012; Bowden et al., 2011; Katsahian et al., 2008; Rondeau et al., 2008; Tudor Smith et al., 2005a; Yamaguchi et al., 2002; Vaida and Xu, 2000; Yamaguchi and Ohashi, 1999; Sargent, 1998; Yau and McGilchrist, 1998). Alternatively, it may be assumed that the baseline hazard functions for each study have the same shape but may have different magnitudes (L5c) (Michiels et al., 2005). Further extensions of these models may consider time-dependent frailties (Yau and McGilchrist, 1998).

Similar to other outcome-type data, the hierarchical Cox proportional hazards regression model is a popular approach for analyzing survival data from multiple studies as it requires no assumptions regarding the distribution of the baseline hazard rate (Michiels et al., 2005; Tudor Smith et al., 2005b, 2005a; Sargent, 1998). This model may account for heterogeneity in baseline hazard (that is, the presence of study effects) by stratifying the baseline hazard by study (L5a). Examples of this second strategy are provided in (Crowther et al., 2012; Bowden et al., 2011; Katsahian et al., 2008; Rondeau et al., 2008; Tudor Smith et al., 2005a; Yamaguchi et al., 2002; Vaida and Xu, 2000; Yamaguchi and Ohashi, 1999). In general, the use of stratified models with random treatment effects (that is, L5a or L5c using normally distributed frailty terms) has been recommended as these models maintain the within-study structure while allowing for heterogeneity in study and treatment effects (Tudor Smith et al., 2005a; Yamaguchi et al., 2002).
Unfortunately, there are significant computational challenges in the fitting of random treatment effects to time-to-event outcome data (Crowther et al., 2012; Fisher et al., 2011; Thompson et al., 2010; Sargent, 1998). For this reason, researchers may choose to approximate the baseline hazard function by defining a piece-wise constant or a spline function (Royston and Parmar, 2002; Vaida and Xu, 2000; Sargent, 1998; Yau and McGilchrist, 1998). Alternatively, it has been shown that the Cox proportional hazards model can be approximated with a Poisson GLMM by splitting follow-up time $t_0$ (and event indicator $y_{ij}$) into $j$ narrow time intervals of fixed length ($L5d$) (Crowther et al., 2012). The parameter $\lambda_j$ then represents the baseline hazard rate during the $j^{th}$ time interval, and $\beta$ is once again the log hazard ratio for the treatment effect. An additional advantage of this approach is that the proportional hazards assumption can be relaxed, for example, by replacing the Poisson with a log-gamma distribution or by pre-specifying the baseline function $h_0(t)$ in $L5b$ (Barrett et al., 2012; Siannis et al., 2010).

3.4.6. Other. Several one-stage models have been proposed for IPD-MA where the outcome belongs to another type of non-normally distributed data (Thompson et al., 2001), recurrent events (Haines and Hill, 2011), repeated measurements (Jones et al., 2009), or multivariate responses (Kim et al., 2013; Riley et al., 2008; Yamaguchi et al., 2002).

3.5. Meta-analysis of individual participant data combined with aggregate data

Subject-level data are often unavailable for all relevant studies (Ravva et al., 2014; Donegan et al., 2013; Saramago et al., 2012; Jansen, 2012; Riley et al., 2008; Sutton et al., 2008; Riley et al., 2007). Therefore, published AD may represent an additional source of evidence. Including published AD in an IPD-MA is strongly recommended when studies providing IPD systematically differ from studies for which IPD were not obtained, as this may lead to data availability bias or reviewer selection bias (Ahmed et al., 2012). But even when there is no indication of such bias, inclusion of AD in an IPD-MA for studies of which IPD are lacking may be considered to increase the statistical power for detecting treatment effects or treatment-covariate interactions (Donegan et al., 2013; Jansen, 2012; Saramago et al., 2012). In general, there are three alternate approaches for combining IPD and AD in a meta-analysis:

- Meta-analysis of AD

In this approach, the available IPD are first reduced to AD (Ravva et al., 2014; Riley et al., 2007) and then pooled into a weighted average as in two-stage meta-analysis. When investigating the presence of effect modification, caution is warranted to avoid the introduction of ecological bias (Signorovitch et al., 2012; Jackson et al., 2008; Riley et al., 2008, 2007; Tudur Smith et al., 2005b; Wakefield and Salway, 2001).

- Meta-analysis of reconstructed IPD

It is possible to reconstruct IPD from published aggregate information for binary (based on $2 \times 2$ tables), ordinal (based on the number of responses within each treatment category), and survival type data (based on Kaplan–Meier survival curves) (Guyot et al., 2012; Riley and Steyerberg, 2010; Riley et al., 2007). In particular, this summary information can be used to restore subject-level information on treatment group and outcome status (or time until event). Afterwards, (traditional) one-stage models can be used for evidence synthesis. Unfortunately, reconstructed IPD do not have subject-level information on covariates and may therefore lead to ecological bias when investigating the presence of effect modification.

- Hierarchical-related regression

This approach directly combines the likelihood from each data source (IPD or AD), for example, using Bernoulli (IPD) and binomial (AD) distributions for modeling the individual responses and trial-specific summaries of binary outcome data (Ravva et al., 2014; Saramago et al., 2012; Jansen, 2012; Sutton et al., 2008; Jackson et al., 2008, 2006; Goldstein et al., 2002). In this manner, it ensures that all studies contribute toward estimation of the overall treatment effect and study-level covariates and that only IPD are used to estimate subject-level covariate effects. Hierarchical related regression is also known as shared parameters models (Donegan et al., 2013) and can, for instance, be achieved by fitting a multilevel model that includes a dummy variable indicating whether a data source relates to IPD or AD (Riley et al., 2008, 2007; Goldstein et al., 2000).

Unfortunately, the incorporation of AD in an IPD-MA does not necessarily improve the validity of estimated treatment effects and interactions. In particular, amongst other issues, the exclusion of subjects post randomization (Tierney and Stewart, 2005) as well as the presence of limited follow-up data and informative drop-out may introduce bias in published AD. Because researchers may often find themselves between the Scylla and Charybdis when deciding to include published AD in an IPD-MA, it has been recommended to assess the sensitivity of meta-analysis results to the inclusion of AD (Riley and Steyerberg, 2010; Riley et al., 2008, 2007). This can, for instance, be achieved by performing an IPD-only meta-analysis and meta-analysis that combines IPD and AD. An additional advantage of sensitivity analyses is that they may help readers without (advanced) statistical expertise to interpret results in function of the underlying assumptions.
3.6. Multiple treatment comparisons

In a different article of this series, we discuss in depth the use of network meta-analysis (NMA) models for synthesizing information from multiple treatments across different studies (Efthimiou et al., 2015). These models allow the estimation of the relative treatment effect between many (rather than two) specific interventions, even when these interventions have never been compared head-to-head. The implementation of NMA models is, however, less straightforward than traditional meta-analysis models and requires studies to have similar covariate distributions for important effect modifiers (Signorovitch et al., 2010). For this reason, it is important to correct for baseline imbalance between studies (Ali et al., 2013; Jansen, 2012) and explore the presence of treatment-covariate interactions (Donegan et al., 2012; Jansen, 2012; Signorovitch et al., 2012). This is ideally achieved using IPD.

The specification of IPD-NMA models can be viewed as an extension of regression model (1) where a dummy variable is included for each (except one) treatment (Higgins et al., 2001). It is, however, more common to describe NMA models as a function of consistency equations, which simplifies the specification of interaction effects (Donegan et al., 2013; Jansen, 2012; Saramago et al., 2012). The use of IPD in NMA has been advocated even if IPD are only available for a fraction of the studies (Jansen, 2012). Recently, several NMA models have been proposed that allow the combined use of IPD and AD in a joint analysis (Donegan et al., 2013; Jansen, 2012; Saramago et al., 2012). These models were shown to have improved network consistency when compared to NMA models that were solely based on AD. Similar to AD-NMA, the quality of evidence resulting from an IPD-NMA strongly depends upon the quality and choice of included studies, the complexity and credibility of statistical models, and the transparency of reporting. A questionnaire for assessing the relevance and credibility of NMA studies has recently been reported by the International Society for Pharmacoeconomics and Outcomes Research and may help to improve the quality of IPD-NMA studies (Jansen et al., 2014).

3.7. Cross-design synthesis

Most trials are conducted to establish efficacy and safety of a single treatment in a specific study setting, and therefore it is often difficult to translate this evidence to “real-world” effectiveness where different concerns may be at play (Prevost et al., 2000; Stuart et al., 2011). For this reason, there is a growing interest to overcome the weaknesses of individual study designs by carefully combining different types of evidence (Droitcour et al., 1993). In particular, intervention research might benefit by the inclusion of non-randomized study data such as from (case-)cohort studies, case-control studies, or patient registries (Higgins et al., 2013; Kaizar, 2011; Peters et al., 2005; Prevost et al., 2000; Li and Begg, 1994). This strategy is more commonly known as cross-design synthesis or generalized synthesis of evidence (Sutton and Higgins, 2008; Sutton et al., 1998).

Because non-randomized studies of interventions (NRSI) are inherently prone to confounding, there is substantial controversy about the magnitude and interpretation of potential differences between the results from RCTs and NRSI (Reeves et al., 2003; Valentine and Thompson, 2013; Schmidt et al., 2013; Golder et al., 2011; Ioannidis et al., 2001; Concato et al., 2000). Whether or not it is valid to combine RCT and NRSI in IPD-MA and under what conditions is an ongoing debate in the literature (Higgins et al., 2013; Reeves et al., 2003). So far, cross-design synthesis has only been advocated when the likelihood of conducting randomized trials is low (e.g., due to ethical concerns or practical limitations) or when research questions relate to unintended, rare (often harms), or long-term outcomes (benefits or harms). The justification of including NRSI may further increase when the derived estimates of relative treatment effects are deemed to be resistant against bias. This becomes more plausible when there are no systematic differences between the groups being compared (e.g., in terms of care provided, exposure to other factors, baseline characteristics, or in terms of how outcomes are determined).

To improve the validity of meta-analyses incorporating evidence resulting from NRSI, several suggestions have been made. First, researchers should critically appraise the risk of bias in available studies prior to cross-design synthesis and carefully decide whether inclusion is appropriate (Berger et al., 2014; Higgins et al., 2013; Reeves et al., 2003; Wells et al., 2013). Subsequently, researchers should routinely account for confounding factors within studies. Finally, researchers should explore potential sources of between-study heterogeneity and compare the results between randomized studies and NRSI (Valentine and Thompson, 2013).

3.8. Dealing with missing data

Missing data are a common problem when analyzing IPD from any (single) study and are typically classified into missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Rubin, 1987). These types correspond to situations where the probability of data being missing is completely independent of observed and unobserved factors (MCAR), depends only on the observed data (MAR), or also depends on unobserved data (MNAR) (Donders et al., 2006; Schafer, 1999; Rubin, 1987). When data are missing within a single study, it is generally recommended to apply multiple imputation strategies by adopting a statistical model that assumes MCAR or MNAR. We can distinguish the following missing data scenarios in an IPD-MA (Sutton et al., 1998):
3.9. Software and estimation techniques for individual participant data meta-analysis

Several software packages exist for conducting an IPD-MA and adopt a frequentist or a Bayesian estimation framework (Tables 2 and 3). Frequentist methods typically maximize the likelihood function of the meta-analysis model using expectation–maximization (Vaida and Xu, 2000), Newton–Raphson, or Fisher scoring algorithms. Unfortunately, maximization of the (log-)likelihood function becomes problematic when there are few studies in the meta-analysis or included studies are small (Crowther et al., 2012; Li et al., 2011; Thompson et al., 2001; Higgins et al., 2001). For this reason, it is often recommended to adopt more advanced estimation techniques such as restricted (or residual) maximum likelihood estimation (Bowden et al., 2011; Higgins et al., 2001), penalized maximum likelihood estimation (Rondeau et al., 2008), penalized quasi-likelihood (Thompson et al., 2001), extended quasi-likelihood estimation, penalized partial likelihood estimation, or semi-parametric-penalized

![Table 2. Software packages for fitting meta-analysis models](image)

<table>
<thead>
<tr>
<th>Software</th>
<th>Package</th>
<th>Characteristics</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>WinBUGS, JAGS, Stan, OpenBUGS</td>
<td>--</td>
<td>Fitting of one-stage and two-stage meta-analysis models using Bayesian Markov chain Monte Carlo (MCMC)</td>
<td>[26, 32, 48, 52, 67, 81, 86, 95, 98, 118, 126, 146, 147]</td>
</tr>
<tr>
<td>NONNEM, R, S-plus*</td>
<td>--</td>
<td>Fitting of GLMM</td>
<td>[87]</td>
</tr>
<tr>
<td>ecoreg</td>
<td>--</td>
<td>Estimation of individual-level covariate-outcome associations using AD (&quot;ecological inference&quot;) or a combination of AD and individual participant data (IPD) (&quot;hierarchical-related regression&quot;)</td>
<td>[52]</td>
</tr>
<tr>
<td>hglm</td>
<td>--</td>
<td>Fitting of GLMM where the random effect may come from a conjugate exponential-family distribution</td>
<td>--</td>
</tr>
<tr>
<td>lme4</td>
<td>--</td>
<td>Fitting of GLMM using ML or REML (for mixed linear models only)</td>
<td>[11, 30, 56, 65, 110, 153], SI3</td>
</tr>
<tr>
<td>MASS</td>
<td>--</td>
<td>Fitting of GLMM using penalized quasi-likelihood (PQL)</td>
<td>--</td>
</tr>
<tr>
<td>mvmeta</td>
<td>--</td>
<td>Meta-analysis and meta-regression of AD (two-stage meta-analysis)</td>
<td>[30], SI3</td>
</tr>
<tr>
<td>nlme</td>
<td>--</td>
<td>Fitting of linear mixed-effects models using ML or REML</td>
<td>[89, 122]</td>
</tr>
<tr>
<td>survival</td>
<td>--</td>
<td>Fitting of Cox PH and mixed effect survival models using penalized partial likelihood estimation (PPL)</td>
<td>[44, 122, 132, 133]</td>
</tr>
<tr>
<td>frailtypack</td>
<td>--</td>
<td>Fitting of frailty models using semi-parametric-penalized likelihood (SPL)</td>
<td>[97]</td>
</tr>
<tr>
<td>coxme</td>
<td>--</td>
<td>Fitting mixed effects Cox PH models</td>
<td>--</td>
</tr>
</tbody>
</table>

GLMM, generalized linear mixed model; AD, aggregate data; PH, proportional hazards; REML, restricted (or residual) maximum likelihood; ML, maximum likelihood.

References are provided in Supporting Information (SI) 4.

*An overview of software packages for two-stage meta-analysis can be found on [http://cran.r-project.org/web/views/MetaAnalysis.html](http://cran.r-project.org/web/views/MetaAnalysis.html).
likelihood estimation. Unfortunately, these techniques are not widely implemented to investigate nonlinear outcomes (e.g., binary, ordinal, or count data).

The Bayesian framework combines the likelihood function with prior information (Yamaguchi et al., 2002; Higgins et al., 2001; Sargent, 1998; Larose and Dey, 1997) and is typically implemented within WinBUGS, JAGS, OpenBUGS, or Stan. The use of prior information may help to overcome convergence issues when studies are small or few studies are available (Table 4). Semi-informative prior distributions may, for instance, be used when model parameters are unidentifiable due to a lack of data (Jackson et al., 2008). Alternatively, informative prior distributions may be used, for example, when relevant information can be borrowed from historical data Table 3.

### Table 3. Software packages for fitting meta-analysis models

<table>
<thead>
<tr>
<th>Software Package</th>
<th>Characteristics</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS – PROC GLIMMIX</td>
<td>Fitting of GLMM using penalized quasi-likelihood (PQL)</td>
<td>[82, 129]</td>
</tr>
<tr>
<td>PROC GLM</td>
<td>Fitting of GLMM using MOM</td>
<td>[148]</td>
</tr>
<tr>
<td>PROC LOGISTIC</td>
<td>Fitting of mixed nonlinear models (binary/ordinal/nominal responses) using ML</td>
<td>[147]</td>
</tr>
<tr>
<td>PROC MIXED</td>
<td>Fitting of mixed linear models using ML, REML, or MOM. Can also perform two-stage meta-analysis.</td>
<td>[48, 57, 93]</td>
</tr>
<tr>
<td>PROC NLMIXED</td>
<td>Fitting of mixed nonlinear models using (approximated) ML</td>
<td>[95, 114, 148]</td>
</tr>
<tr>
<td>Stata – gllamm</td>
<td>Fitting of GLMM using ML</td>
<td>–</td>
</tr>
<tr>
<td>mvmeta</td>
<td>Meta-analysis and meta-regression of aggregate data (AD) (two-stage meta-analysis)</td>
<td>[5]</td>
</tr>
<tr>
<td>REGOPROB2</td>
<td>Fitting of random effects generalized-ordered probit models</td>
<td>[26, 95, 122]</td>
</tr>
<tr>
<td>stmixed</td>
<td>Fitting of flexible parametric survival models with mixed effects</td>
<td>–</td>
</tr>
<tr>
<td>XT</td>
<td>Fitting of GLMM</td>
<td>–</td>
</tr>
<tr>
<td>MLwiN, MLn, FORTRAN, C</td>
<td>Fitting of GLMM and survival models using ML, REML, and EM</td>
<td>[41, 48, 122, 126, 136, 147]</td>
</tr>
</tbody>
</table>

GLMM, generalized linear mixed model; REML, restricted (or residual) maximum likelihood; ML, maximum likelihood; MOM, method of moments; EM, expectation maximization.

References are provided in Supporting Information 4.

Table 4. Overview of prior distributions used within the Bayesian framework

<table>
<thead>
<tr>
<th>Parameters for fixed effects (regression coefficients)</th>
<th>95% prior belief</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-informative β ∼ N(0, 10^4)</td>
<td>[0.00; +∞) for the odds ratio</td>
<td>[98, 118]</td>
</tr>
<tr>
<td>β ∼ N(0, 10^5)</td>
<td>[0.00; +∞) for the odds ratio</td>
<td>[26]</td>
</tr>
<tr>
<td>β ∼ N(0, 10^6)</td>
<td>[0.00; +∞) for the odds ratio</td>
<td>[32, 48, 70, 86, 95, 126, 147]</td>
</tr>
<tr>
<td>β ∼ N(0, 10)</td>
<td>[0.00; 494] for the odds ratio</td>
<td>[86]</td>
</tr>
<tr>
<td>Weakly informative β ∼ N(0, 1.47)</td>
<td>[0.20; 5.03] for the odds ratio</td>
<td>[52]</td>
</tr>
</tbody>
</table>

Parameters for between-study standard deviation

<table>
<thead>
<tr>
<th>Parameters for between-study standard deviation</th>
<th>95% prior belief</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-informative τ^2 ∼</td>
<td>[1.2 × 10^7, +∞) for τ</td>
<td>[67]</td>
</tr>
<tr>
<td>τ^2 ∼ U(0.01, 100)</td>
<td>[1.62; 9.87] for τ</td>
<td>[70]</td>
</tr>
<tr>
<td>τ ∼ N(0.10, 0)</td>
<td>[0.30; 22.48] for τ</td>
<td>[146]</td>
</tr>
<tr>
<td>τ ∼ U(0, 10)</td>
<td>[0.25; 9.76] for τ</td>
<td>[32, 118]</td>
</tr>
<tr>
<td>Weakly informative τ ∼ N(0.5)</td>
<td>[0.12; 4.88] for τ</td>
<td>[86]</td>
</tr>
<tr>
<td>τ ∼ U(0, 2)</td>
<td>[0.05; 1.96] for τ</td>
<td>[98]</td>
</tr>
<tr>
<td>τ ∼ N(0, 1)(0)</td>
<td>[0.03; 2.24] for τ</td>
<td>[26, 95]</td>
</tr>
<tr>
<td>τ ∼ U(0, 1)</td>
<td>[0.02; 0.98] for τ</td>
<td>[126]</td>
</tr>
<tr>
<td>τ ∼ N(0, 0.125)(0)</td>
<td>[0.01; 0.80] for τ</td>
<td>[86]</td>
</tr>
<tr>
<td>τ^2 ∼ U(1 - 10^-3)</td>
<td>[0.05; 0.62] for τ</td>
<td>[52]</td>
</tr>
<tr>
<td>τ ∼ N(0, 0.033)(0)</td>
<td>[0.01; 0.41] for τ</td>
<td>[86]</td>
</tr>
</tbody>
</table>

References are provided in Supporting Information 4.
An additional advantage of the Bayesian framework is that it allows more flexibility in choosing and combining likelihood functions that are appropriate for the data that are being analyzed. As such, the use of Bayesian meta-analyses may sometimes help to improve the credibility of estimated treatment effects and between-study heterogeneity. The Bayesian framework, however, requires careful specification of prior distributions, investigation of convergence, and the drawing of inferences from a large simulated sample. Because it tends to be cumbersome and requires substantial statistical expertise, Bayesian IPD-MA is not pursued by many researchers.

4. Concluding remarks

Individual participant data meta-analyses offer numerous advantages over meta-analyses that are solely based on published AD and are therefore considered as a gold standard in evidence synthesis. Nevertheless, the presence of publication bias, data accessibility bias, or reviewer selection bias may still hamper or invalidate the results of an IPD-MA (Ahmed et al., 2012), and synchronizing between different sources of IPD is not always straightforward. For this reason, it is not recommended to conduct IPD-MA on an ad hoc basis (e.g., without systematic review). Because the implementation of an IPD-MA requires additional efforts and statistical expertise, researchers should carefully assess whether the potential advantages (e.g., increased power, reduced bias, and investigating interaction and subgroup effects) outweigh the extra efforts involved. Finally, researchers should be aware that IPD-MA may have similar issues to meta-analyses of published AD and are no panacea against poorly designed and conducted primary research.

Acknowledgements

The research leading to these results was conducted as part of the GetReal consortium. For further information, please refer to www.imigetreal.eu.

Funding

The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115546, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/20072013) and EFPIA companies’ in-kind contribution.

References


Haines TP, Hill AM 2011. Inconsistent results in meta-analyses for the prevention of falls are found between study-level data and patient-level data. *Journal of Clinical Epidemiology* **64**: 154–162. DOI:10.1016/j.jclinepi.2010.04.024.


**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s web site.