Meta-analysis of individual participant data: rationale, conduct, and reporting

Richard D Riley,[1] Paul C Lambert,[2] Ghada Abo-Zaid[3]

[1]Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham B15 2TT

[2]Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester LE1 7RH

[3]School of Mathematical Sciences, University of Birmingham, Birmingham B15 2TT

Correspondence to: R D Riley
r.d.riley@bham.ac.uk

The use of individual participant data instead of aggregate data in meta-analyses has many potential advantages, both statistically and clinically. **Richard D Riley and colleagues** describe the rationale for an individual participant data meta-analysis and outline how to conduct this type of study

Meta-analysis methods involve combining and analysing quantitative evidence from related studies to produce results based on a whole body of research. As such, meta-analyses are an integral part of evidence based medicine. Traditional methods for meta-analysis synthesise aggregate study level data obtained from study publications or study authors, such as a treatment effect estimate (for example, an odds ratio) and its associated uncertainty (for example, a standard error or confidence interval). An alternative but increasingly popular approach is meta-analysis of individual participant data, or individual patient data, in which the raw individual level data for each study are obtained and used for synthesis.[1] In this article we describe the rationale for individual participant data meta-analysis and illustrate through applied examples why this strategy offers numerous advantages, both clinically and statistically, over the aggregate data approach.[1 2] We outline when and how to initiate an individual participant data meta-analysis, the statistical issues in conducting one, how the findings should be reported, and what challenges this approach may bring.

What are individual participant data?

The term "individual participant data" relates to the data recorded for each participant in a study. In a hypertension trial, for example, the individual participant data could be the pre-treatment and post-treatment blood pressure, a treatment group indicator, and important baseline clinical characteristics such as age and sex, for each patient in each study (table). A set of individual participant data from multiple studies often comprises thousands of patients; this is the case in the table, so for brevity we do not show all rows of data here.

This concept is in contrast to the term "aggregate data," which relates to information averaged or estimated across all individuals in a study, such as the mean treatment effect on blood pressure, the mean age, or the proportion of participants who are male. Such aggregate data are derived from the individual participant data themselves, so individual participant data can be considered the original source material.

What is an individual participant data meta-analysis?

As with any meta-analysis, an individual participant data meta-analysis aims to summarise the evidence on a particular clinical question from multiple related studies, such as whether a treatment is effective. The statistical implementation of an individual participant data meta-analysis crucially must preserve the clustering of patients within studies; it is inappropriate to simply analyse individual participant data as if they all came from a single study. Clusters can be retained during analysis by using a two step or a one step approach.[3] In the two step approach, the individual participant data are first analysed in each separate study independently by using a statistical method appropriate for the type of data being analysed; for example, a linear regression model might be fitted for continuous responses such as blood pressure, or Cox regression might be applied for time to event data such as mortality. This step produces aggregate data for each study, such as a mean treatment effect estimate and its standard error. These data are then synthesised in the second step using a suitable model for meta-analysis of aggregate data, such as one that weights studies by the inverse of the variance while assuming fixed or ran-

SUMMARY POINTS

Meta-analysis of individual participant data involves obtaining and then synthesising the raw individual level data from multiple related studies

The application of individual participant data meta-analysis has risen dramatically from just a few articles a year in the early 1990s to an average of 49 a year since 2005

The use of individual participant data for meta-analysis has both statistical and clinical advantages; for example, it increases the power to detect differential treatment effects across individuals in randomised trials and allows adjustment for confounding factors in observational studies

Individual participant data meta-analyses should be protocol based, clearly reported, driven by clinical questions, and used when a meta-analysis of (published) aggregate data cannot reliably answer the clinical questions

Statistical methods for meta-analysis of individual participant data must preserve the clustering of patients within studies. Either a one step or a two step approach should be used

Individual participant data meta-analyses are often resource intensive but can be facilitated by the collaboration of research groups

Example of individual participant data from 10 hypertension trials that assess effect of treatment versus placebo on systolic blood pressure

Study ID	Patient ID	Age (years)	Sex (1=male, 0=female)	Treatment group (1=treatment, 0=control)	Systolic blood pressure before treatment (mm Hg)	Systolic blood pressure after treatment (mm Hg)
1	1	46	1	1	137	111
1	2	35	1	0	143	133
...	...	...	...	...	...	...
1	1520	62	0	0	209	219
2	1	55	0	1	170	155
2	2	38	1	1	144	139
...	...	...	...	...	...	...
2	368	44	1	0	153	129
3	1	51	1	1	186	166
3	2	39	0	1	201	144
...	...	...	...	...	...	...
3	671	54	0	0	166	141
...	...	...	...	...	...	...
10	1	71	0	1	149	128
10	2	59	1	0	168	169
...	...	...	...	...	...	...
10	978	63	0	1	174	128

Dotted line indicates where non-displayed rows of data occur.

Hypothetical data based on Wang et al.[27]

dom (treatment) effects across studies. In the one step approach, the individual participant data from all studies are modelled simultaneously while accounting for the clustering of participants within studies. This approach again requires a model specific to the type of data being synthesised, alongside appropriate specification of the assumptions of the meta-analysis (for example, of fixed or random effects across studies).

Detailed statistical articles regarding the implementation and merits of one step and two step individual participant data meta-analysis methods are available.[4-10] The two approaches have been shown to give very similar results, particularly when the meta-analysis aims to estimate a single treatment effect of interest.[9 11 12] One step individual participant data meta-analyses conveniently require only a single model to be specified, but this may increase complexity for non-statisticians and requires careful separation of within study and between study variability.[9] Two step individual participant data meta-analyses are clearly more laborious, but in the second step they allow the use of traditional, well known meta-analysis techniques
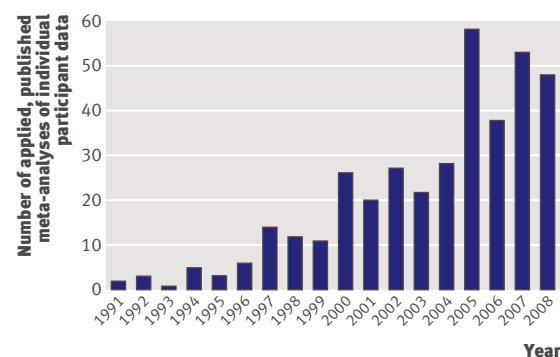
Fig 1 | Number of distinct, applied meta-analyses of individual participant data published up to March 2009,* as identified by a systematic review of Medline, Embase, and the Cochrane Library. *Six articles published in 2009 were identified up to 5 March, when the review was conducted

such as those used by the Cochrane Collaboration[13] (for example, inverse variance fixed effect or random effects approach, or the Mantel-Haenszel method).[14] Importantly, both one step and two step approaches produce results that can inform evidence based practice, such as a pooled estimate of treatment effect across studies and how the treatment effect is modified by study level characteristics (for example, dose of treatment or study location) and patient level characteristics (for example, age or stage of disease).

Incidence of individual participant data meta-analyses over time

To assess the changes in the publication frequency of applied articles using an individual participant data meta-analysis, we performed a systematic review of the published literature. We searched Medline, Embase, and the Cochrane Library up to March 2009 using a set of search terms as described elsewhere.[15] We defined an "applied individual participant data meta-analysis" article as one describing the application and findings of a meta-analysis of individual participant data from multiple healthcare studies or multiple collaborating research groups. There was no restriction on the type of studies being synthesised. Methodological articles, commentaries, or discussion articles regarding individual participant data meta-analysis were not included.

Our review identified 383 distinct, applied individual participant data meta-analysis articles published up to March 2009 (fig 1).[15] Only 57 articles (15%) were published before 2000, after which there was a considerable rise in the number of published articles, with an average of 49 articles published a year between 2005 and 2009. This growth is most likely the result of an increased awareness of why individual participant data meta-analyses are beneficial and the initiation of collaborations of research groups specifically to perform such studies. The 383 articles focused predominately on cancer, cardiovascular disease, and diabetes, and most assessed whether a treatment or intervention was effective, often in subgroups of patients. The assessment of risk factors for disease onset or prognostic factors for disease outcome was also popular, being the primary aim in 86 (22%) of the 383 articles, which signifies a recognition that individual participant data are particularly advantageous for time to event analyses.

When do an aggregate data meta-analysis and an individual participant data meta-analysis coincide?

In an "aggregate data meta-analysis," researchers try to replicate the two step approach to meta-analysis of individual participant data. In the first step, suitable aggregate data are extracted from the study publications or obtained directly from the study authors to allow the second step to be implemented. For example, a treatment effect estimate and its standard error are sought from each study for synthesis. If the required aggregate data can be obtained, a two step individual participant data meta-analysis and an aggregate data meta-analysis will be equivalent, provided other factors are equal (for example, number of patients, follow-up length, and so on).[8 9 12] Indi-

Box 1|: Potential advantages of using individual participant data for a meta-analysis

- Consistent inclusion and exclusion criteria can be used across studies, and, if appropriate, individuals who were originally excluded can be reinstated into the analysis
- Missing data can be observed and accounted for at the individual level
- Results presented in the original study publications will be verified (assuming individual participant data provided can be matched to the individual participant data used in the original analyses)
- Up to date follow-up information, potentially longer than that used in the original study publications, may be available
- Studies that contain the same or overlapping sets of participants can be identified
- Results for missing or poorly reported outcomes can be calculated and incorporated, thus reducing the problem of selective within study reporting
- Results for unpublished studies can be calculated and incorporated, thus reducing the problem of publication bias
- The statistical analysis can be standardised across studies (for example, the analysis method, how continuous variables are analysed, the time points assessed, and so on) and more appropriate or advanced methods can be applied where necessary
- Model assumptions in each study can be assessed, such as proportional hazards in Cox regression models, and complex relationships like time dependent effects can be modelled
- Estimates adjusted for baseline (prognostic) factors can be produced where previously only unadjusted estimates were available, which may increase statistical power and allow adjustment for potential confounding factors
- Baseline (prognostic) factors can be adjusted for consistently across studies
- Meta-analysis results for specific subgroups of participants can be obtained across studies (for example, those receiving a particular treatment or those with a particular biomarker level), and differential (treatment) effects can be assessed across individuals, which can help reduce between study heterogeneity
- Prognostic models (risk scores) can be generated and validated, and multiple individual level factors can be examined in combination (for example, multiple biomarkers and genetic factors, and their interaction)
- The correlation between multiple end points can be accounted for when each participant provides results at multiple time points, as in a meta-analysis of longitudinal data

vidual participant data are not needed if all the required aggregate data can be obtained in full from authors or the published papers themselves, which may save considerable time and resources. Researchers should note that the required aggregate data to extract will depend on the clinical questions and on the outcomes of interest and the most appropriate statistical measures to assess them. Thus liaison between clinicians and statisticians is crucial to identify the aggregate data needed. Also a scoping exercise, or evidence from a previous similar review, may help establish whether such aggregate data are obtainable and reliable.

What are the advantages of a meta-analysis of individual participant data?

Meta-analysis of individual participant data has many potential advantages, both statistically and clinically, over meta-analysis of aggregate data (box 1). Aggregate data are often 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, amplifying the threat of publication bias and within study selective reporting. On the contrary, having individual participant data facilitates standardisation of analyses across studies and direct derivation of the information desired, independent of significance or how it was reported. Individual participant data may

also have a longer follow-up time, more participants, and more outcomes than were considered in the original study publication. This means that individual participant data meta-analyses are potentially more reliable than aggregate data meta-analyses, and the two approaches may lead to different conclusions. Four applied examples that illustrate this are shown below.[9] [16-18]

Differences in conclusions with regard to a treatment effect
Example 1: Effectiveness of laparoscopic repair at reducing persistent pain
A two step meta-analysis of individual participant data from hernia surgery trials showed that laparoscopic repair significantly reduced persistent pain compared with open repair (odds ratio 0.54, 95% CI 0.46 to 0.64).[16] In contrast, an earlier meta-analysis of published aggregate data indicated a statistically significant benefit in favour of open repair (odds ratio 2.03, 95% CI 1.03 to 4.01). This disparity occurred because the individual participant data analysis included an additional 17 trials, as few useable published aggregate data were available owing to outcome reporting bias. Furthermore, reanalysis of the individual participant data from one trial produced markedly different results to its published aggregate data.

Example 2: Effectiveness of paternal white blood cell immunisation at reducing recurrent miscarriage
The effect of paternal white blood cell immunisation on reducing recurrent miscarriage was greater in a meta-analysis of aggregate data from four published articles (live birth rate ratio (RR) 1.29, 95% CI 1.03 to 1.60) than in a two step meta-analysis of individual participant data provided by the same investigators (RR 1.17, 95% CI 0.97 to 1.37) or in a two step meta-analysis of individual participant data from four other unpublished trials (RR 1.01, 95% CI 0.74 to 1.28).[17] The effect of immunisation was statistically significant (P<0.05) only in the aggregate data meta-analysis. The different conclusions were the result of additional patients in the analysis of individual participant data and publication bias affecting the analysis of published aggregate data.

Differences in conclusions with regard to how patient level characteristics modify treatment effect
Example 1: Effect of elevated panel reactive antibodies on the effectiveness of antilymphocyte antibody induction
An individual participant data meta-analysis of five randomised trials of antilymphocyte antibody induction therapy for renal transplant recipients showed that treatment was significantly more effective among patients with elevated (20% or more) panel reactive antibodies than in those without elevated panel reactive antibodies.[18] The estimated difference in the odds of treatment failure between patients who didn't have elevated levels and those who did was −1.33 (P=0.01) on the $\log_e$ scale. This important difference in treatment effect was not detected by an aggregate data meta-analysis (estimated difference in $\log_e$ odds −0.014; P=0.68), because it had far less power to detect differential treatment effects across individuals than did the individual participant data analysis.
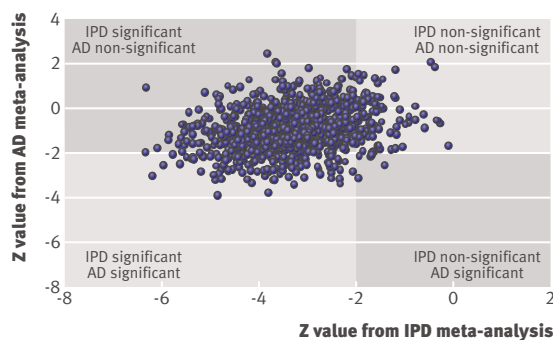
Fig 2 |Comparison of the power of meta-analyses to detect a differential treatment effect across two groups of patients when individual participant data (IPD) or aggregate data (AD) are used. Adapted from Lambert et al[19]

Fig 3 |An example of ecological bias within an aggregate data meta-analysis

This point is illustrated in figure 2, which considers a simulated set of 1000 meta-analyses, each containing five trials where the treatment was effective for high risk patients but ineffective for low risk patients.[19] The statistical power of the meta-analysis technique used is given by the proportion of the 1000 meta-analyses that correctly detect this differential treatment effect with statistical significance (that is, the proportion that give a Z value of less than 1.96, which equates to a P value of <0.05 and a differential treatment effect estimate in favour of high risk patients). The power is much higher when using individual participant data (90.8%) than when using aggregate data (14.8%). This difference is because individual participant data meta-analyses can model the individual risk status across (often hundreds of) patients; in contrast, aggregate data meta-analyses can only model the proportion of high risk patients across (usually a few) studies.

Example 2: Effect of gender on the effectiveness of hypertension treatment
In an aggregate data meta-analysis of 10 hypertension trials,[9] the across trials interaction between the proportion of male participants and the mean treatment effect on systolic blood pressure was estimated as 15.10 mm Hg (95% CI 8.78 to 21.41), indicating that the treatment effect was significantly less in men than in women. However, a one step individual participant data meta-analysis revealed that the within trials difference in treatment effect between men and women was actually 0.89 (95% CI 0.07 to 1.30, which is statistically but not clinically significant.[9]

This issue is shown graphically in figure 3. Each block represents a trial and the block size is inversely proportional to the standard error of the trial's treatment effect estimate. The fitted dashed lines are from the individual participant data analysis and their gradient denotes the difference in treatment effect (in mm Hg) between men and women in each trial. These lines are quite flat and suggest little difference between men and women. In contrast, the gradient of the fitted solid line across trials from the aggregate data analysis suggests a much larger treatment effect for women compared with men. The bias here in the aggregate data analysis is known as ecological bias,[18] and is probably caused by confounding across trials. The individual participant data meta-analysis is more
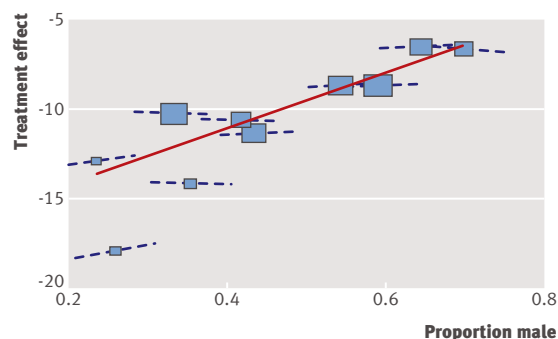
reliable as it directly assesses patient level information and so avoids the problem of trial level confounding

Beyond the "grand mean"
The previous two examples importantly show that an individual participant data meta-analysis can produce more clinically relevant results than a meta-analysis of aggregate data, going beyond the "grand mean" toward individualised medicine.[20] For example, subgroups of patients with a common characteristic (such as those who are female) can be identified within individual participant data and thus a meta-analysis can be conducted specifically for that group, with increased power compared with the individual studies themselves (fig 2). In contrast, the aggregate data approach is reliant on extracting summary results for each subgroup, which will often not be available. Similarly, analysis of individual participant data allows more powerful and reliable examination of treatment effects across individuals,[18 20] as within trial information can be directly used to estimate how patients' characteristics modify treatment benefit (fig 3).[21] Multiple individual level factors can also be examined together, such as the predictive ability of combinations of biomarkers and genetic factors,[22] the results of which may facilitate stratified medicine.[23] Analyses can even be adjusted for baseline (prognostic) factors, which can increase the power of the analysis to detect a true treatment effect[24] and allows adjustment for confounding factors (a particular advantage for analyses of observational studies).

Individual participant data also enable the statistical modelling of complex relationships, such as non-linear trends and time dependent treatment effects. They even facilitate the development of prognostic models (or risk scores),[25] as a model can be constructed using individual participant data from some of the studies and then its performance validated using the individual participant data from remaining studies.[26] It is thus very clear why individual participant data are deemed the "gold standard" for meta-analysis.[1]

What are the disadvantages of a meta-analysis of individual participant data?
Meta-analysis of individual participant data is not without its disadvantages. In particular, this approach is resource intensive, because substantial time and costs are required to contact study authors, obtain their individual partici-

pant data, input and "clean" the provided individual participant data, resolve any data issues through dialog with the data providers, and generate a consistent data format across studies. For example, Ioannidis et al[22] undertook an individual participant data meta-analysis in 2002 that required 2088 hours for data management, with 1000 emails exchanged between study collaborators and the data managers. The required costs and time will clearly vary depending on the complexity of the analysis and the number of studies involved,[27] and such factors need serious consideration before embarking on an individual participant data meta-analysis or when applying for grant income. In particular, resource requirements must be considered for both the team conducting the individual participant data meta-analysis and the original study authors themselves. The latter group is often neglected, but cooperation of the original study authors is crucial to the success of the project and these individuals will often commit many hours "cleaning" and updating their data and resolving ongoing queries.

The individual participant data approach may also require advanced statistical expertise, and there may be ethical or confidentiality concerns about using patient level data. In our experience, individual participant data meta-analyses usually have the same objectives as the original studies, for which ethical approval should exist; however, this must be verified and, if in doubt, advice or approval must be sought from an ethics committee. When asking for individual participant data, researchers should stipulate that individuals' names and contact details must be erased from the data before they are supplied, so that participants can be identified only via a unique ID number interpretable solely by the original study authors.

Although a high proportion of the desired individual participant data can usually be obtained,[15] sometimes data are not available for all studies. Individual participant data might have been be lost or destroyed, or study authors may not be contactable or willing to collaborate. An individual participant data meta-analysis may then be biased if the provision of individual participant data is associated with the study results. In such a situation, it is important to examine any differences between studies that provided individual participant data and studies that did not provide individual participant data, and, if possible, consider whether the conclusions of the meta-analysis might change if those studies not providing individual participant data had been included. Meta-analysis methods for combining individual participant data with aggregate data are available for this purpose.[9][15] In many situations, however, such as the assessment of differential treatment effects across individuals,[9] aggregate data will add very little to an individual data meta-analysis and serve only to amplify why individual participant data was desired.

It is also important to recognise that the quality of individual participant data is dependent on the quality of the original studies themselves, and that the studies providing data are not necessarily of the highest quality. A meta-analysis of individual participant data from a set of poorly designed trials with many potential sources of bias is as deficient as a meta-analysis of aggregate data from these trials. Individual participant data meta-analyses should thus also include a quality assessment of the original studies and, if appropriate, make clear how the inclusion of lesser quality studies impacts on conclusions. If only low quality studies exist, it may be better to initiate a prospective individual participant data meta-analysis.

How to obtain individual participant data for a meta-analysis

Two crucial steps toward undertaking an individual participant data meta-analysis are deciding how much individual participant data are needed and obtaining the individual participant data themselves. One option is to adopt a systematic review approach, where all relevant published and unpublished studies are identified through a transparent, systematic search, and then study authors are contacted to provide the individual participant data. Authors may be more willing to agree if the clinical and methodological reasons for requiring the individual participant data are clearly outlined, preferably via a face to face meeting. It may also help to promise regular updates on the results of the individual participant data meta-analysis and provide the incentive of joint authorship on subsequent publications. Patience, politeness, and good communication are essential.

Another option is to collaborate with other research groups and agree to pool resources to answer specific clinical questions. For example, members of the receptor and biomarker group within the European Organisation for Research and Treatment of Cancer provided 18 datasets for an individual participant data meta-analysis of prognostic markers in breast cancer.[28] Such meta-analyses can also be updated prospectively as new data become available from collaborators. One concern, however, is that studies within the collaboration may not reflect the entire set of existing studies, potentially introducing bias. It is thus important for collaborations to be inclusive. For example, one individual participant data meta-analysis by the Early Breast Cancer Trialists' Collaborative Group involved over 400 named collaborators,[29] who commendably provided individual participant data for 42 000 women from 78 randomised treatment comparisons.

Reporting individual participant data meta-analyses

A review of 33 applied individual participant data meta-analyses from between 1999 and 2001 noted that "clear reporting of the statistical methods used was rare" and that only a few studies actually referred to a protocol for their individual participant data project.[3] Clearly these shortcomings must be addressed. Like all good research, meta-analyses of individual participant data should be protocol driven and conducted with clear and prespecified objectives. Studies should also be clearly reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[30] or Meta-analysis Of Observational Studies in Epidemiology (MOOSE)[31] guidelines for meta-analyses of randomised controlled trials and observational studies, respectively. Given that these guidelines are not specific to the individual participant data approach, in box 2 we outline some additional information that individual participant data meta-analyses should report.

Box 2 |: Suggested information to report from an individual participant data meta-analysis, to supplement those reporting guidelines of PRISMA and MOOSE[30 31]

- Whether there was a protocol for the individual participant data project, and where it can be found
- Whether ethics approval was necessary and (if appropriate) granted
- Why the individual participant data approach was initiated
- The process used to identify relevant studies for the meta-analysis
- How authors of relevant studies were approached for individual participant data
- How many authors (or collaborating groups) were approached for individual participant data, and the proportion that provided such data
- The number of authors who did not provide individual participant data, the reasons why, and the number of patients (and events) in the respective study
- Whether those authors who provided individual participant data gave all their data or only a proportion; if the latter, then describe what information was omitted and why
- Whether there were any qualitative or quantitative differences between those studies providing individual participant data and those studies not providing individual participant data (if appropriate)
- The number of patients within each of the original studies and, if appropriate, the number of events
- Details of any missing individual level data within the available individual participant data for each study, and how this was handled within the meta-analyses performed
- Details and reasons for including (or excluding) patients who were originally excluded (or included) by the source study investigators
- Whether a one step or a two step individual participant data meta-analysis was performed, and the statistical details thereof, including how clustering of patients within studies was accounted for
- How many patients from each study were used in each meta-analysis performed
- Whether the assumptions of the statistical models were validated (for example, proportional hazards) within each study
- Whether the individual participant data results for each study were comparable with the published results, and, if not, why not (for example, individual participant data contained updated or modified information)
- How individual participant data and non-individual participant data studies were analysed together (if appropriate).
- The robustness of the meta-analysis results following the inclusion or exclusion of non-individual participant data studies (if appropriate)

An applied example of an individual participant data meta-analysis of hypertension trials

An individual participant data meta-analysis of 10 hypertension trials was conducted with the objective of estimating the effect of a treatment on systolic blood pressure.[9 32] The applied one step and two step individual participant data meta-analysis models used are described fully elsewhere.[9] Briefly, they involve fitting linear regression models that estimate the treatment effect on systolic blood pressure, having adjusted for systolic blood pressure at baseline. This "analysis of covariance" approach is the most appropriate method for analysing the change from baseline in a continuous variable[33] because it gives the most precise and least biased estimate of treatment effect. It is very difficult to undertake an analysis of covariance meta-analysis without individual participant data because individual studies analyse and report a change from baseline in a heterogeneous fashion.[34]

The one step and two step individual participant data meta-analyses gave identical results. The pooled treatment effect was estimated at −10.16 (95% CI −12.27 to −8.06), indicating that hypertension treatment reduced systolic blood pressure by, on average, 10.16 mm Hg compared with controls. There was also, however, a large between study variance of 7.13, indicating that the treatment effect varied considerably across the trials. A 95% prediction interval[35] for the underlying treatment effect in a new trial was estimated as −16.69 to −3.63. This range indicates that although there is heterogeneity among trials, in any single trial hypertension treatment is effective at reducing systolic blood pressure.

Availability of individual participant data in this example also allowed a reliable and powerful assessment of how patient level characteristics modify the treatment effect. For example, a one step individual participant data analysis showed no clinically important difference in treatment effect between men and women. In contrast, if only aggregate data had been available, the analysis would have wrongly indicated a clinically important treatment effect difference in favour of women (fig 3).[9] A one step individual participant data analysis also revealed a non-linear effect of age on the treatment effect. Up to the age of 55 years, the treatment effect increased for each year increase in age; however, after 55 years there was no evidence of differential treatment effects according to age.[9] This finding was not detectable without individual participant data.

Conclusions

The decision to undertake an individual participant data meta-analysis should be driven by the clinical questions of interest and whether a meta-analysis of (published) aggregate data can reliably answer them. In many situations an individual participant data meta-analysis will offer considerable advantages, both statistically and clinically, over a meta-analysis of aggregate data, which is why individual participant data approaches are increasingly being applied.

Important challenges remain, however, not least the task of obtaining the individual participant data itself. Detailed commentaries exist regarding the often laborious and expensive process of retrieving and processing individual participant data.[22 36] Ways of addressing this difficulty include storing individual participant data in a central repository or on the internet,[37] but perhaps most crucial is the initiation of collaborations across research groups in each field. Even when individual participant data are fully available, obstacles may still remain because studies are usually collated retrospectively. Completed studies may be of poor quality (for example, poorly designed with selection biases), may not have recorded important variables, or may have used outdated treatment strategies.[38] For such reasons prospectively planned individual participant data meta-analyses have been advocated. This technique is similar to undertaking a multicentre trial, except it allows variations in the protocols of included studies. This approach maximises the power of the meta-analysis by achieving consistency in, for example, treatments received, outcomes assessed, variables collected, and how data are recorded. Prospectively planned meta-analyses are achievable[39] but currently few have been completed, so it would be encouraging to see growth in the use of this method.

1   Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418-22.
2   Oxman AD, Clarke MJ, Stewart LA. From science to practice. Meta-analyses using individual patient data are needed. *JAMA* 1995;274:845-6.
3   Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005;2:209-17.
4   Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000;19:3417-32.
5   Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001;20:2219-41.
6   Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Stat Med* 2001;20:2243-60.
7   Tudur-Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 2005;24(9):1307-19.
8   Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials* 2009;6:16-27.
9   Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008;27:1870-93.
10   Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;27:6111-36.
11   Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 1998;54:317-22.
12   Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics* 1999;55:1221-3.
13   Higgins JPT, Green S. *Cochrane handbook for systematic reviews ofinterventions version 5.0.1* [updated September 2008]. The Cochrane Collaboration, 2008. www.cochrane-handbook.org.
14   Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719-48.
15   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;60:431-9.
16   McCormack K, Grant A, Scott N. Value of updating a systematic review in surgery using individual patient data. *Br J Surg* 2004;91:495-9.
17   Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data. Paternal cell immunization for recurrent miscarriage. *JAMA* 1995;274:830-6.
18   Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21:371-87.
19   Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;55:86-94.
20   Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? *BMJ* 1997;315:1610-4.
21   Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 2005;365:341-6.
22   Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR. Meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002;156:204-10.
23   Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov* 2007:287-93.
24   Hernandez AV, Eijkemans MJ, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? *Ann Epidemiol* 2006;16:41-8.
25   Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
26   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;23:907-26.
27   Stewart LA, Clarke MJ for the Cochrane Working Group. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995;14:2057-79.
28   Look MP, van Putten WL, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* 2002;94:116-28.
29   Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;366:2087-106.
30   Moher D, Liberati A, Tetzlaff J, Altman DG for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
31   Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283:2008-12.
32   Wang JG, Staessen JA, Franklin SS, Fagard R, Gueyffier F. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension* 2005;45:907-13.
33   Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow-up measurements. *BMJ* 2001;323:1123-4.
34   Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. *Stat Med* 2005;24:3823-44.
35   Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A* 2009;172:137-59.
36   Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002;25:76-97.
37   Hutchon DJ. Publishing raw data and real time statistical analysis on e-journals. *BMJ* 2001;322:530.
38   Chia S, Bryce C, Gelmon K. The 2000 EBCTCG overview: a widening gap. *Lancet* 2005;365:1665-6.
39   Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 2005;366:1267-78.