# Survey finds that most meta-analysts do not attempt to collect individual patient data

**Stephanie A. Kovalchik, Ph.D.**
Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd., EPS 8047, Rockville, MD 20852

## Abstract

**Objective**—To characterize current efforts and outcomes of individual patient data (IPD) collection among meta-analysts of randomized controlled clinical trials.

**Study Design and Setting**—Corresponding authors of meta-analyses of randomized controlled trials in general medicine with a binary endpoint were sent an e-mail survey inquiring about their efforts to obtain IPD. Descriptive statistics of each meta-analysis were extracted to evaluate their association with data-seeking.

**Results**—Only 22 (4.2%) of the sampled meta-analyses included IPD. Out of 360 authors surveyed, 256 (71%) reported not seeking IPD: 48% thought the undertaking would be too difficult, 30% thought it was not necessary for their main analysis, 25% did not have sufficient time or resources, and 22% never considered it. Seeking IPD was not significantly associated with any trial characteristic examined, including whether subgroup analyses were performed.

Authors who sought IPD obtained a median of 2 datasets (IQR = 0 to 5). Unsuccessful contact (43%), refusal without explanation (21%), and lost or inaccessible data (20%) were the most common reasons why trial data could not be obtained.

**Conclusion**—The infrequency of attempts made by meta-analysts to obtain participant data is an important contributor to the rarity of IPD meta-analyses.

### Keywords

effect modification; individual patient data; meta-analysis; meta-regression; randomized controlled trial; subgroup analysis

## Introduction

Treatment effect modification is a fundamental concern of clinical research (1). Though subgroup analyses are an important strategy for identifying the sources of heterogeneity in treatment response, they can lack credibility when based on data from a single clinical trial (2). For this reason, there is increasing interest in using meta-analysis to improve the validity of assessments of subgroup effects (3).

The author has no conflicts of interest to declare.

In contrast to aggregate data meta-analyses that are based on summaries of study-level statistics, an individual patient data (IPD) meta-analysis is an analysis of pooled participant-level data from all of the relevant primary studies. The principal advantages of an IPD meta-analysis are increased power and protection against ecological bias in assessments of heterogeneity of treatment effect (4–6). Other potential benefits include outcome and methodological harmonization, the ability to conduct time-to-event analyses, and collaboration with the primary study investigators (7).

Despite these advantages, challenges to collecting patient-level trial data could make the implementation of IPD meta-analyses infeasible for most researchers. The barriers to IPD meta-analysis have been frequently cited (8–10), yet their extent and nature are not well understood because they have not been studied empirically. The present article reports findings from an investigation to characterize the current frequency and relative importance of barriers to seeking and obtaining patient-level data as part of a meta-analysis in the clinical sciences.

## Methods

A search of MEDLINE identified patient-level and aggregate data meta-analyses published between May 2010 and January 2011. Studies included in the analytic sample were English language meta-analyses of parallel group randomized controlled trials whose primary analysis was based on a binary patient outcome. The reason for focusing on analyses of the same type of outcome was in order to assess methodological characteristics of the reviews under more uniform conditions. When there were multiple reviews from the same corresponding author, one review was randomly selected for inclusion. Event counts and sample sizes were extracted for each treatment group of the primary analysis. When the primary analysis was not explicitly designated, the first reported analysis of a binary outcome was used. The reporting of subgroup analyses was also recorded.

Corresponding authors were sent a brief questionnaire asking about their efforts to obtain patient-level data from the trials included in their primary analysis (Supplemental Material, File S1). Authors were asked to provide the number of trial investigators contacted for IPD and to describe the responses received. When no trialists were contacted, authors were asked to list all of the reasons why IPD collection was not undertaken. Univariate logistic regression analyses were performed to determine whether the decision to seek IPD was associated with the meta-analysis' number of trials, total sample size, treatment effect size, between-trial heterogeneity (measured by Higgins $I^2$ (11)), or reporting of subgroup effects. In these analyses, the response variable was whether or not the study had sought participant-level data.

## Results

From 1194 results of the initial MEDLINE search, abstracts were read until a sample of 500 meta-analyses were identified that met the study inclusion criteria (Supplemental Material, File S2). The reviews represented a diverse set of medical disciplines published in 272 different journals; 45 (9%) from the Cochrane Database of Systematic Reviews and 11 (2.2%) from the British Medical Journal. The meta-analyses included a median of 6 trials (IQR = 4 to 10), 185 subjects per trial, and the mean odds ratio of treatment effect was 0.71 (IQR = 0.51 to 0.87). Most (87.2%) summarized treatment effect with the DerSimonian-Laird random-effects model (12); only 22 (4.2%) of the meta-analyses included IPD. At least one subgroup analysis was performed in 60.6% of the reviews, most frequently using the Q partition test or meta-regression (95%) (13–14).

A multivariate logistic regression was performed to compare the trial characteristics of studies whose corresponding authors did and did not respond to the email survey. The odds of being a responding trial was significantly associated with reporting an IPD meta-analysis (18 vs. 3 trial, adjusted OR 4.8 95% CI = 1.6 to 20.8) but was not statistically significantly associated with the pooled sample size, number of primary studies, or main treatment effect.

Of the 360 (72%) corresponding authors who completed the survey, 273 (76%) reported that they had not attempted to collect patient-level data (Table 1). Most authors did not collect IPD because they were doubtful of success (48%). For 14% of these authors, this belief came after an initial effort to gather IPD had been unsuccessful. Nearly 30% of the authors reported that they did not believe there was a statistical advantage to using IPD for their primary analysis, though 64% of these authors also performed subgroup analyses. 23% of the authors had never considered collecting IPD, while 22% did not believe they had sufficient time or resources to do so.

The decision to seek IPD was statistically significantly associated with the number of authors on the study (OR 1.13 95% = 1.04 to 1.25) but was not significantly associated with any of the remaining meta-analysis characteristics examined (Table 2). This suggests that factors external to the study analysis had a greater influence on the choice to undertake a patient-level meta-analysis. The finding that the odds of pursuing an IPD meta-analysis increased with a greater number of authors could reflect the greater resources required to complete a meta-analysis of participant data. This interpretation is supported by a statistically significant mean difference of 6 authors (P-value < 0.001) that was found between published IPD meta-analyses and non-IPD meta-analyses.

Authors who sought IPD (n = 104) reported being able to obtain a median of 2 IPD datasets out of a median of 4 solicited trials (Table 3). The within-review ratio of the number of IPD datasets obtained to the number sought was a median of 50%. In terms of the pooled sample size, the median percentage of the total sample size obtained was 65% (IQR = 15.6 to 100%). Nearly one-quarter of authors who attempted to collect IPD obtained no patient-level datasets. The most frequent reasons for an unsuccessful solicitation were non-response (43.0%), refusal without explanation (21.2%), lost or inaccessible data (20.1%), and sponsor prohibition (5.6%). Explicit concerns about publishing competition or data security were uncommon.

## Discussion

Despite the increase in IPD meta-analyses since the mid-1990s, meta-analyses with patient-level data remain rare. Fewer than 5% of this study's sample of meta-analyses of randomized controlled trials in general medicine used IPD, confirming the prevalence previously proposed by Simmonds et al. (14). From a survey of the efforts of a large sample of meta-analysts, the present study has helped to shed light on the relative importance of barriers to performing IPD meta-analyses. While actual and perceived unavailability of participant data were both key contributors, perceived unavailability was the single most important factor. Approximately half of the 71% of meta-analyses that did not undertake an IPD meta-analysis were owing to the principal investigators' belief that participant data would not be attainable.

The findings from the study's survey indicate that this common perception might be overly pessimistic. On average, meta-analysts who requested IPD for the purpose of a quantitative review obtained datasets for 50% of the eligible trials and 65% of the total possible number of participants, on average. These are lower than the success rates found by Riley, Simmonds and Look (15) and Simmonds et al. (14), who reported 64% for the average

percentage of IPD trials obtained and 80% for the total participants obtained, respectively. However, these figures are not directly comparable to those reported here owing to differences in study selection criteria. In the reviews of Riley, Simmonds and Look and Simmonds et al., the samples were based on published meta-analyses of patient-level data. As a consequence, the success rates they report might over-estimate the rates for the larger population of meta-analyses that have attempted but not necessarily published an IPD meta-analysis.

Although the present study included both types of meta-analyses, there are several reasons why the availability rates reported here might have still over-estimated the actual frequency of obtaining IPD data that was sought for the purpose of performing a meta-analysis. The present study's success rates did not include 17 meta-analyses that had initiated but later abandoned an effort to collect IPD, since the corresponding authors of these studies were generally unable to provide sufficient enough information about their collection efforts to be analyzed. Also, the higher proportion of published IPD meta-analyses among the survey respondents as compared to the non-respondents raises the possibility of a non-response bias that resulted in inflated availability rates.

It is surprising that more meta-analysts do not seek participant data given the number of advantages that an IPD meta-analysis offers. IPD is needed to conduct survival analyses, to ensure uniformity of outcomes and methods, and to protect against bias and low power when exploring the causes of treatment effect heterogeneity (16–18). The infrequency of IPD meta-analysis has often been attributed to the labor involved (19–20). Yet compared to the 48% of the surveyed meta-analysts who did not try to collect IPD because they doubted whether it could be obtained, only 25% cited lack of resources as a reason. This suggests that concern over a wasted effort is a greater barrier to undertaking an IPD meta-analysis than the resource-intensiveness of the effort itself.

Almost half of the non-seeking meta-analysts reported that they either had never considered performing an IPD meta-analysis or they did not believe that doing so would offer any statistical advantage. This statistic is even more notable when one considers that 64% of these authors had performed at least one subgroup analysis. A conclusion to take from this is that a concerning portion of the meta-analytic community does not appreciate the value of participant data for quantitative reviews.

The results of the present study raise several important questions. In order to encourage a high response rate, the study questionnaire was designed to be brief. This prevented the collection of more detailed data on the characteristics of the meta-analytic study and the IPD collection process, such as the number of attempts made and the time and resources expended. Given the high percentage of meta-analysts who were found to have made no attempt to collect IPD, it would be useful for future research to elaborate on the underlying reasons for this. In order to have a more uniform sample, the study was restricted to meta-analyses of a primary binary endpoint. Consequently, it could not be determined whether the IPD collection characteristics of meta-analyses of continuous or time-to-event outcomes differ from the sample considered. Also, the focus of the present study was limited to the requestors of IPD. Thus, the perspective of trialists and, specifically, their reasons for sharing and not sharing patient-level trial data, remains unclear.

Publishers of systematic reviews and meta-analyses could have a vital role in addressing these remaining issues. A description of the decision to collect or not to collect IPD and the results of any attempted efforts should be made a standard part of the reporting of quantitative reviews. Encouraging meta-analysts to routinely report this information could

have the dual benefits of promoting the collection of IPD and adding to our knowledge about on-going barriers to the successful completion of participant-level meta-analysis.

## Conclusions

More than 70% of meta-analysts make no attempt to gather patient level data for their quantitative review. The most frequently reported reason is the perception that IPD will not be available, a perception that is not supported by observed success rates among meta-analysts who have sought IPD.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

1. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med. 2007; 357(21):2189–2194. [PubMed: 18032770]

2. Rothwell PM. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. Lancet. 2005; 365:176–186. [PubMed: 15639301]

3. Fisher DJ, Copas AJ, Tierney JF, Parmar MK. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol. 2011; 64(9):949–976. [PubMed: 21411280]

4. Simmonds MC, Higgins JPT. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007; 26(15):2982–2999. [PubMed: 17195960]

5. 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(3):371–387. [PubMed: 11813224]

6. 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(1):86–94. [PubMed: 11781126]

7. van Walraven C. Individual patient meta-analysis-rewards and challenges. J Clin Epidemiol. 2010; 63:235–237. [PubMed: 19595573]

8. Vickers AJ. Whose data set is it anyway? Sharing raw data from randomized trials. Trials. 2006; 7:15. [PubMed: 16704733]

9. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol. 2005; 5:14. [PubMed: 15850485]

10. 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(1):76–97. [PubMed: 11868447]

11. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327(7414):557–560. [PubMed: 12958120]

12. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–188. [PubMed: 3802833]

13. Deeks, JJ.; Higgins, JP.; Altman, DG. Analysing and presenting results. In: Alderson, P.; Green, S.; Higgins, JPT., editors. Cochrane Reviewer's Handbook 4.2.1. The Cochrane Library. Chichester: John Wiley; 2004. p. 68-139.

14. Simmonds MC, Higgins JPT, Stewert 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; 2:209–217. [PubMed: 16279144]

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(5):431–439. [PubMed: 17419953]

16. Riley RD, Lambert PC. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010; 340:c221. [PubMed: 20139215]

17. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level factors. J Clin Epidemiol. 2004; 57:683–697. [PubMed: 15358396]

18. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PloS journals. PloS One. 2009; 4(9):e7078. [PubMed: 19763261]

19. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? Lancet. 1993; 341:418–422. [PubMed: 8094183]

20. Clarke MJ, Stewart LA. Meta-analyses using individual patient data. J Eval Clin Pract. 1997; 3:207–212. [PubMed: 9406108]

## Table 1

Reasons why authors of meta-analyses of randomized controlled trials did not collect individual patient data (n=256).

| Reason[*] | No. (%)[†] |
|---|---|
| Doubtful of success | 123 (48.0) |
| Abandoned effort after initial attempt | 17 (13.8) |
| No statistical advantage | 76 (29.7) |
| Insufficient time or resources | 65 (25.4) |
| Was never considered | 58 (22.7) |
| Lacked expertise | 12 (4.7) |
| Update of meta-analysis of aggregate data | 9 (3.5) |
| Concern of bias if IPD is not available for some trials | 4 (1.6) |
| Avoid institutional review | 2 (0.1) |

Abbreviations: IPD, individual patient data

[*] Respondents could provide multiple reasons.

[†] Percentage of authors.

**Table 2**

Study factors associated with seeking patient-level data (n=360).

| Study Factor[*] | OR[†] | 95% CI |
|---|---|---|
| Number of trials | 1.01 | (0.99, 1.03) |
| Total patient sample size (log-scale) | 0.98 | (0.81, 1.16) |
| Treatment effect[‡] | 2.00 | (0.61, 6.94) |
| Higgins' $I^2$ | 1.00 | (0.98, 1.01) |
| Subgroup analysis | 0.82 | (0.48, 1.40) |
| Number of authors | 1.13 | (1.04, 1.25) |

[*] Factors were based on primary analysis of a binary outcome.

[†] Univariate logistic regression analysis.

[‡] The reciprocal was taken for odds ratios > 1.

**Table 3**

Data sharing characteristics reported by authors of meta-analyses of randomized controlled trials who sought individual patient data (n=104).

| Characteristic | No. (%)[*] |
|---|---|
| Number of trialists contacted for IPD, median (IQR) | 4 (2, 7) |
| Number of trials for which IPD was obtained, median (IQR) | 2 (0, 5) |
| Attempts resulting in no IPD (of 104) | 25 (24) |
| Reasons IPD was not obtained [†] | |
| Unable to reach author, no response | 77 (43.0) |
| Refusal without explanation | 38 (21.2) |
| Data was lost or inaccessible | 36 (20.1) |
| Sponsor prohibits data sharing | 10 (5.6) |
| Unwilling to collaborate | 7 (3.9) |
| Still publishing | 6 (3.4) |
| Offered data but did not follow-up | 3 (1.7) |
| Available data was incompatible | 1 (0.6) |
| Security concerns | 1 (0.6) |

Abbreviations: IPD, individual patient data; IQR, interquartile range

[*] Percentage of all solicited trials unless indicated otherwise.

[†] Respondents could provide multiple reasons.