Retrieval of Individual Participant Data for a Meta-Analysis

Running Head: Retrieval of Individual Participant Data

Type of Article: Short Communication

Abstract Word Count: 192

Manuscript Word Count (Introduction, Methods, Results, Discussion, Conclusions: 1519

Corresponding Author: Dr. George A. Kelley\textsuperscript{a}, FACSM, Professor and Director, Meta-Analytic Research Group, \textsuperscript{a}School of Public Health, Department of Biostatistics, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, WV 26506-9190, Office Phone: 304-293-6279, Fax: 304-293-5891, E-mail: gkelley@hsc.wvu.edu

Co-Author: Kristi S. Kelley\textsuperscript{a}, M.Ed., Research Instructor, \textsuperscript{a}School of Public Health, Department of Biostatistics, Robert C. Byrd Health Sciences Center, West Virginia University, PO Box 9190, Morgantown, WV 26506-9190, Office Phone: 304-293-6280, Fax: 304-293-5891, E-mail: kskelley@hsc.wvu.edu
Abstract

**Purpose:** Despite the potential for individual participant data (IPD) meta-analysis to yield more valid results than aggregate data (AD) meta-analysis, the feasibility of obtaining IPD is not well established. The purpose of this study was to add to that literature. **Methods:** Using data from a previous meta-analysis of 29 studies on exercise in adults with arthritis and other rheumatic disease, the percentage of studies in which useable IPD was provided was calculated. Exact logistic regression was used to examine the association between the retrieval of IPD with year of publication and country in which the study was conducted (USA versus Other) included as potential predictors. A two-tailed alpha value $< 0.05$ was considered statistically significant.

**Results:** Eight of 29 authors (27.6%, 95% CI = 11.3% to 43.8%) provided useable IPD. Neither year of publication (odds ratio = 1.05, 95% CI = 0.90 to 1.27, $p = 0.58$) nor country (odds ratio = 1.36, 95% CI = 0.20 to 10.9, $p = 1.00$) was significantly associated with the obtainment of IPD. **Conclusions:** Aggregate data meta-analysis may be preferable to IPD meta-analysis. However, further research is warranted before any definitive recommendations can be made.

**Key words:** meta-analysis; methods; aggregate data; individual participant data; individual patient data; systematic review

**Abbreviations**

IPD – Individual participant data

AD – aggregate data
1. Introduction

The prevalence of meta-analyses has increased substantially over approximately the past 25 years. For example, a recent PubMed search by the first author on February 21, 2015 using the keyword “meta-analysis” found that the number of citations increased from 331 in 1990 to 14,329 in 2014 (unpublished results). While aggregate data (AD) meta-analysis, an approach in which summary statistics from eligible studies are pooled, is still the most common type of meta-analysis, individual participant data (IPD) meta-analysis pools the results from different studies based on data from each participant [1]. Two of the potential advantages of IPD meta-analyses are the ability to examine covariates at the individual level as well as a potential reduction in reporting and publication biases [1-4]. However, a major disadvantage may be the ability to retrieve IPD from eligible studies, thus leading to a smaller number of included trials and potentially biased results on the topic of interest. The investigative team has previously reported the retrieval of IPD from only 29 of 76 (38.2%) eligible studies dealing with the effects of exercise on bone mineral density in adults [5]. However, this study was conducted approximately 13 years ago. Since that time, technological advances have improved one’s ability to store and share data, thereby making it easier to share de-identified IPD with others. In addition, knowledge regarding the feasibility of obtaining IPD, an important factor when making decisions about whether an AD or IPD meta-analysis should be conducted, is not well established [6-8]. The purpose of this short communication was to add to this literature.

2. Materials and Methods
2.1. Data source. Data were derived from a recently published AD meta-analysis in which an IPD meta-analysis was originally planned, details of which have been described elsewhere [9]. Briefly, studies were included if they were randomized controlled trials examining the effects of exercise (aerobic, strength training, or both) on depressive symptoms in adults with arthritis and other rheumatic diseases [9]. Twenty-nine studies representing 2,449 participants (1,470 exercise, 979 control) were included [10-38].

2.2. Retrieval of IPD. Using a pre-defined form letter (see Supplementary File 1), de-identified IPD was requested by having the second author contact the corresponding author of each eligible study via electronic mail asking if they would be interested in providing IPD. A response was requested within two weeks, regardless of interest, with the authors being informed that they would be listed in the acknowledgements section of each published study derived from the project if they provided their IPD. If no response was received within two weeks, up to two additional requests were sent via electronic mail. For those who responded but chose not to contribute, reasons given for not participating, if any, were recorded. For those authors who expressed interest in providing IPD, a second electronic mail was sent that included an attachment consisting of a pre-defined list of IPD requested (see Supplementary File 2). Investigators were asked to provide IPD in a format that was convenient for them within four weeks. If IPD was not received within four weeks, as many as two additional reminders were sent via electronic mail. The dates of all communications were recorded.

2.3. Statistical Analysis. Descriptive statistics were used to describe the number of responses to initial electronic mail requests for IPD, number of days to respond to initial
electronic mail requests, number of authors who provided useable IPD, and number of days from initial requests to receipt of IPD. Reasons for not providing IPD from authors who were willing to supply such information were also recorded. Furthermore, descriptive statistics were calculated for the two potential predictors included in the regression model.

Because of the small sample size, exact logistic regression was used to examine for potential predictors with respect to whether or not IPD was received [39]. Based on previous research [5], the two potential predictors included in the model were country in which the study was conducted and year that the study was published. The chi-square distribution ($\chi^2$) was used to examine the overall model. The alpha level for statistical significance was set at $p \leq 0.05$.

3. Results

3.1. Descriptive Statistics for Retrieval of IPD. The authors from 20 of 29 studies (69.0%, 95% CI = 52.2% to 85.8%) responded to initial electronic mail requests for IPD while 9 (31.0%, 95% CI = 14.2% to 47.8%) never responded despite multiple requests. The response time to initial requests varied widely from 1 to 181 days ($\bar{X} \pm SD = 54.2 \pm 74.8$, 95% CI = 21.4 to 87.0, Mdn = 17). Eight of 29 authors (27.6%, 95% CI = 11.3% to 43.8%) provided useable IPD. The number of days from initial requests for data to receipt of IPD ranged from 36 to 179 ($\bar{X} \pm SD = 74.4 \pm 46.4$, 95% CI = 42.0 to 106.8, Mdn = 64). Reasons for not providing IPD included not having the data any longer (n=4) and time (n=1). Another author said they would supply IPD if a consortium was formed and in which they were included as a co-author. Year of publication ranged from 1989 to 2011. Thirteen studies (44.8%, 95% CI = 26.7% to 62.9%) were conducted in the United
States while the remaining 16 (55.2%, 95% CI = 37.1% to 73.3%) were conducted in other countries than the United States.

3.2. Potential Predictors in the Obtainment of IPD. Results for exact logistic regression are shown in Table 1. The overall model was not statistically significant ($\chi^2 = 0.62$, $p=0.71$) and neither year of publication nor country were significant predictors for the receipt of IPD ($p>0.05$ for both).

4. Discussion

4.1. Overall Findings. The current study suggests that the obtainment of IPD was low, with less than one third of eligible studies providing such. The inability to obtain IPD from the majority of eligible studies may bias results and limit one from conducting analyses based on individual versus grouped data, one of the very reasons for conducting an IPD meta-analysis [1,4]. While methods exist for combining IPD and AD [8,40], the investigative team was not comfortable using those methods because of the inability to obtain the data necessary to replicate the results reported in this previous work (Dr. Richard Riley, personal electronic mail communication, July 11, 2012, Dr. Jan Staessen, personal electronic mail communication, July 30, 2012) [40].

The response rates observed are either similar to [5,6], lower [7,8] or higher [41] than previous research. Most notably, this latter study was only able to retrieve IPD for 15% of eligible participants [41]. Possible reasons for the lower response rates observed in the current study include the lack of a consortium, larger number of eligible studies and older publication years for eligible studies. The lack of association between year and country with the obtainment of IPD are in contrast with previous research where a trend was found for both to be associated with the retrieval of IPD [5]. One possible reason
for this discrepancy may have been the smaller number of studies included (29 versus 76) [5].

Finally, as previously reported [9], no statistically significant or clinically important differences were found in depressive symptoms between those studies that supplied IPD versus those that did not, a finding consistent with previous research [5].

4.2. Implications for Research and Practice. Given that this was a case study, a need exists for additional work that includes multiple IPD meta-analyses. This should include data regarding the time and costs involved in conducting an IPD meta-analysis, something that was beyond the scope of the current investigation.

Given the inability to obtain the majority of IPD, an AD meta-analysis may be preferable. However, if one chooses to conduct an IPD meta-analysis, the length of time to retrieve IPD needs to be considered. In addition, while the development of a consortium between authors of the original studies may increase the amount of IPD obtained, this should be balanced with the additional time involved. Most notably, any potential benefit of an IPD meta-analysis should be considered with respect to the increased costs. For example, Steinberg et al., estimated that the costs associated with conducting a meta-analysis of 12 studies was more than 5 times greater using the IPD versus AD approach [42] while others estimated the costs of this same study to be at least 8 times greater given that the investigative team continued to work on the study after funding for the project ended [43].

4.3. Strengths and Potential Limitations Strengths. To the best of the authors’ knowledge, this is one of the most recent studies regarding the feasibility of retrieving IPD. However, since the study focused on one attempted IPD meta-analysis, the
findings may not be generalizable to other IPD meta-analyses. In addition, no cost data were collected or analyzed, thereby limiting the applicability of results. Finally, the collapsing of countries other than the US into one category because of the small number of studies available for each country could have biased the results.

5. Conclusions

The results of the current study suggest that an AD meta-analysis may be preferable to an IPD meta-analysis. However, further research is warranted before any definitive recommendations can be made.

Acknowledgments and Funding

The authors would like to thank the following people for supplying us with IPD: Susan Buckelew, Jennifer L. Etnier, Marlene Fransen, A. Hakkinen, Geri B. Neuberger, Sheila C. O’Reilly, Michael Doherty, Joanna Romawski, Mike E. Miller, Brenda Penninx, William B. Applegate, Candace L. Schachter, Ayse Karan, Sureyya Sencan, Jamil Natour, Valerie Valim, and H. Valkeinen.

This work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (grant number RO1 AR061346) and the National Institute of General Medical Sciences of the National Institutes of Health (grant number U54 GM104942). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest Statement

The Authors declare that there is no conflict of interest.
References

[1] Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? The Lancet 1993;341:418-22.


[25] O'Reilly SC, Muir KR, Doherty M. Effectiveness of home exercise on pain and
disability from osteoarthritis of the knee: a randomised controlled trial. Ann


comparison of aerobic and resistance exercise effects on emotional and physical
function in older persons with high and low depressive symptomatology. J

combination self-management in women with fibromyalgia: a randomized trial.

on key health outcomes in women with fibromyalgia: a randomized controlled

aerobic exercise in sedentary women with fibromyalgia: a randomized controlled

[31] Sencan S, Ak S, Karan A, et al. A study to compare the therapeutic efficacy of
aerobic exercise and paroxetine in fibromyalgia syndrome. J Back Musculoskelet
13


Table 1. Results for exact logistic regression for receipt of IPD (n = 29).

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>SE</th>
<th>p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>1.05</td>
<td>0.09</td>
<td>0.58</td>
<td>0.90, 1.27</td>
</tr>
<tr>
<td>Country</td>
<td>1.36</td>
<td>1.12</td>
<td>1.00</td>
<td>0.20, 10.90</td>
</tr>
</tbody>
</table>

Notes: IPD, Individual participant data; OR, odds ratio; SE, standard error; p, alpha value, calculated as 2 * the probability of the sufficiency statistic, a statistic derived from single-parameter tests of the null hypothesis that the coefficient equals zero versus a 2-sided alternative; 95% CI, 95% confidence interval; Alpha (p) and 95% CI calculated from exact conditional distributions; both independent variables (year and country) calculated separately with the other variable conditioned out of the calculation.