

ORIGINAL ARTICLE

# A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial

Joshua R. Polanin<sup>a,\*</sup>, Mary Terzian<sup>b</sup>

<sup>a</sup>Research & Evaluation, American Institutes for Research, Washington, DC 20007, USA

<sup>b</sup>Development Services Group, Inc, Bethesda, MD 20814, USA

Accepted 9 October 2018; Published online 19 October 2018

## Abstract

**Background and Objectives:** Sharing individual participant data (IPD) among researchers, on request, is an ethical and responsible practice. Despite numerous calls for this practice to be standard, however, research indicates that primary study authors are often unwilling to share IPD, even for use in a meta-analysis. This study sought to examine researchers' reservations about data sharing and to evaluate the impact of sending a data-sharing agreement on researchers' attitudes toward sharing IPD.

**Methods:** To investigate these questions, we conducted a randomized controlled trial in conjunction with a Web-based survey. We searched for and invited primary study authors of studies included in recent meta-analyses. We emailed more than 1,200 individuals, and 247 participated. The survey asked individuals about their transparent research practices, general concerns about sharing data, attitudes toward sharing data for inclusion in a meta-analysis, and concerns about sharing data in the context of a meta-analysis. We hypothesized that participants who were randomly assigned to receive a data-sharing agreement would be more willing to share their primary study's IPD.

**Results:** Results indicated that participants who received a data-sharing agreement were more willing to share their data set, compared with control participants, even after controlling for demographics and pretest values ( $d = 0.65$ , 95% CI [0.39, 0.90]). A member of the control group is 24 percent more likely to share her data set should she receive the data-sharing agreement.

**Conclusions:** These findings shed light on data-sharing practices, attitudes, and concerns and can be used to inform future meta-analysis projects seeking to collect IPD, as well as the field at large. © 2018 Elsevier Inc. All rights reserved.

**Keywords:** Data sharing; Randomized controlled trial; Meta-analysis; Research transparency; Individual participant data

Sharing individual participant data (IPD) among researchers, on request, is an ethical and responsible research practice [1]. Data sharing at the individual participant level (i.e., providing a data set to a colleague or researcher, when requested) has the potential to increase the reproducibility of findings, decrease outcome reporting and publication bias, and bolster the partnerships among academic researchers [2–4]. Despite these positive implications, however, research continues to indicate that primary study authors are unwilling to share IPD with meta-analysts [5–8].

Researchers' unwillingness to share IPD for use in a meta-analysis has the potential to limit the applicability and utility of the IPD meta-analyses, decrease statistical

power, and halt IPD meta-analytic production altogether [9]. It is therefore imperative that meta-analysts attempt to address potential data-sharing concerns when attempting to acquire IPD from primary study authors. One potential means to decrease concerns is through the use of a data-sharing agreement [2]. A data-sharing agreement is a nonbinding document sent to the primary study author with the first email request for the IPD, and it provides information about how the IPD will be interpreted, used, and stored. Empirical information on the effects of a data-sharing agreement on researcher attitudes and behaviors has yet to be collected.

The purpose of this project, therefore, is to investigate the impacts of a data-sharing agreement, offered in the context of a hypothetical data request for use in a meta-analysis, on participants' attitudes toward sharing their data. We also sought to develop a better understanding of participants' data-sharing concerns and use these findings to make modifications to a data-sharing agreement for use in future meta-analytic research. The long-term goal of this

The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the University of California-Berkeley.

\* Corresponding author. American Institutes for Research, 1000 Thomas Jefferson Street NW, Washington, DC 20007, USA. Tel.: +1 202 403 5509; fax: +1 855 459 6213.

E-mail address: [jpolanin@air.org](mailto:jpolanin@air.org) (J.R. Polanin).

**What is new?****Key findings**

- Using a randomized controlled trial design, participants randomly assigned to receive a data-sharing agreement were statistically significantly more likely to be willing to share their data set, compared to those who did not receive the data-sharing agreement.

**What this adds to what was known?**

- Data-sharing attitudes can be improved by preemptively addressing researchers' concerns.

**What is the implication and what should change now?**

- Meta-analysts who communicate with primary study authors should send a data-sharing agreement along with the email asking for the data set.

research is to promote greater use of transparent research practices and increased collaboration between researchers.

### 1. The use of individual participant data for meta-analysis

IPD meta-analyses can offer three distinct advantages over a traditional aggregate meta-analysis [10]. First, IPD meta-analyses decrease potential ecological biases, thus increasing the validity of the results. An ecological bias occurs when an aggregate outcome measure does not accurately depict or represent individual-level phenomena. IPD meta-analyses reduce the likelihood of bias by allowing for relationships that occur at the individual level to be examined. Second, IPD meta-analyses, because of their nature, have greater statistical power compared with traditional meta-analyses [11], making them more likely to detect statistically significant effects. The results from meta-analytic models utilize more observations and thus increase statistical precision relative to traditional aggregate meta-analyses. Third, IPD meta-analyses can accommodate subgroup analyses that are not available when using traditional aggregate meta-analyses [12]. Subgroup analyses, for example, can elucidate intervention effectiveness for particular groups of individuals that may not be seen when viewed in aggregate.

### 2. Obtaining IPD for meta-analysis

To collect IPD for a meta-analysis, meta-analysts may search online repositories for applicable data sets, correspond with a clinical trial when dealing with pharmaceutical data [13], or contact primary study authors directly via email.

Although direct contact with the primary study author has become less popular in pharmaceutical trials, email correspondence with primary study authors remains the primary mode of contact for other industries, especially the social sciences, yet little information is available to guide these interactions. Polanin and Williams [2] suggested sending an email that introduces the project and the project's purpose, outlines how the data will be used, and details what will be done with the data set after the meta-analytic project ends. The authors concluded by suggesting that a data-sharing agreement be sent to the primary study authors, along with the email, to provide this information in a succinct manner.

IPD meta-analysts face significant challenges, however, when attempting to collect IPD directly from primary study authors. A survey by Campbell et al. [14] found that 47 percent of academic faculty who asked other faculty for additional study information or for access to data sets were denied. Indeed, two recent reviews of IPD retrieval seem to confirm these results. Nevitt et al. [15] conducted a systematic review of IPD studies in epilepsy. Although overall the percentage of studies found was moderate, IPD for 80% of eligible studies was available for 49% of total IPD studies; the authors stated that the most common reasons for not collecting IPD were due to the lack of e-mail response or an unwillingness to share the data set. Polanin [7] also conducted a systematic review of IPD studies that focused specifically on studies where emailing the primary author directly constituted the primary method of collecting IPD. Of the 2,745 primary studies the authors attempted to collect, IPD analysts found only 1,599 data set (58%).

Given the importance of IPD to meta-analysis, and the problems that occur while attempting to collect it, meta-analysts have much to gain by studying data-sharing attitudes and practices among primary study authors. Moreover, increasing the rate at which primary study authors share IPD with meta-analysts has the potential to improve meta-analysis' usefulness while increasing research transparency and openness.

As such, the purpose of this project is driven by three research questions: (1) Are participants' attitudes toward sharing IPD increased when they receive a data-sharing agreement, compared with participants who do not receive a data-sharing agreement? (2) What are the reasons why researchers adopt nontransparent research practices (i.e., concerns about sharing IPD)? and (3) What are researchers' attitudes toward and experiences with data sharing in general?

### 3. Methods

This study was registered on Open Science Framework. The study's protocol, preanalysis plan, statistical scripts, and data sets can be found here (<https://osf.io/nkqzu/>). We did not deviate significantly from the original preanalysis plan.

### 3.1. Participants

Primary study authors whose studies were included in a recent meta-analysis constituted the eligible population of participants. We, therefore, first located eligible meta-analyses by searching PubMed, PsycInfo, ERIC, and ProQuest. For the meta-analysis to be included in the sample, it must have met the following criteria: 1) published on or after 2012; 2) focused on a social-science subject area, including psychology, criminology, education, or social work; 3) included at least 20 primary studies; and (4) identified the studies' primary authors. A total of 157 meta-analyses were identified and used. Most meta-analyses examined outcomes related to mental health (61), followed by substance use (25), education (15), criminal justice (13), health and safety (4), and child welfare (1).

From the references of these publications, we extracted 2,239 primary studies and created a database of corresponding authors' names and emails. The primary requirement for inclusion in the sample was that the primary study was published after 2004 and analyzed as part of the meta-analysis. After removing duplicate contacts, we obtained unique email addresses for 1,694 primary study authors. After deleting bounce-back e-mails and inactive accounts, we contacted 1,207 individuals to request their participation in the study. Of these, 580 (48.1%) were randomized to the intervention condition, and 627 (51.9%) were randomized to the control condition. We sent identical recruitment e-mails to intervention and control group participants, changing only the URL associated with the survey condition. Participation was incentivized by offering a chance to win one of five \$100 Amazon gift cards. We used SurveyMonkey to host the survey and collect information.

### 3.2. Design

We received Institutional Review Board approval before conducting the study. Participants were provided an informed consent after entering the online survey but before answering questions. We randomly assigned participants to condition using a random-number generator. Random assignment occurred before the e-mail was sent asking for their consent to participate in the study.

### 3.3. Hypothetical data request conditions

After answering questions from the pretest portion of the survey, we stated the following:

“We would like you to imagine that a researcher conducting a meta-analysis has sent you an email. In the email, the researcher requests that you share the deidentified research data set associated with the study we originally referenced in our email to you. The researcher will include the data set in a meta-analysis.”

The “original e-mail” referred to in the passage refers to the recruitment e-mail we sent to invite participants into the study. In addition to collecting the participants' name and

e-mail address from the original primary study, we collected the title of the primary study included in the meta-analysis and reported this information to them in the e-mail. After reading the passage, the participants “received” the data-sharing agreement or the standard data request, depending on which condition we randomly assigned them to.

#### 3.3.1. Data-sharing agreement condition

In previous research conducted by the first author, a data-sharing agreement had been hypothesized to bolster data-sharing attitudes and behaviors [2]. This one-page document was adapted for use in this study. The agreement is designed to address specific concerns that primary study authors might have on receiving a request to share their IPD.

The data-sharing agreement included three primary sections (see [Online Appendix A](#) for the full document). The first section provided the meta-analyst's name, project title, primary study title, meta-analysis purpose, and reason for requesting the primary study data set. The second section of the data-sharing agreement outlined the responsibilities of the Principal Investigator (PI) and project research team and the rights of the primary study authors (PSAs). The third section included information about the IPD from the PSA, as well as a place for the meta-analyst to provide a signature.

More specifically, statements in the second section of the agreement indicated that the data set would be used solely for current and future meta-analyses conducted by the PI and would not be used to conduct new primary analyses (“Use”); that the PI and research team would have sole access to the data set and would comply with Institutional Review Board requests (“Privacy”); that the data set would be stored in a safe and secure manner (“Security and Storage”); and that the PSA would have the rights to ask for the results of the meta-analysis before publication (“Rights”). The authors developed these statements through personal experience collecting IPD for meta-analysis and through conversations held with experts in the meta-analytic field [2]. Although the data-sharing agreement is not a legal document and therefore nonbinding, we hypothesized that providing this agreement as part of a data request for IPD would improve the data-sharing attitudes of primary study authors who received it.

#### 3.3.2. Standard data request condition

Individuals assigned to the control group “received” the standard data request. In contrast to the data-sharing agreement, the standard data request, developed specifically for this study, asks only if participants would be willing (hypothetically) to share their deidentified IPD for the purpose of a meta-analysis. No other information or agreements were provided to the primary study authors.

### 3.4. Data and procedures

The primary data source for this project was participants' responses to a 30-item online survey, developed by the first and second authors. The survey started with a brief section asking for participants' gender, race/ethnicity, level of education, discipline, and type of employer. All demographic questions were mandatory and answered by all included participants. Next, participants answered questions relating to past research experiences, concerns about data sharing, willingness to share their data, and concerns about sharing their data for use in a meta-analysis.

#### 3.4.1. Transparent research practices

The pretest portion of the online survey included seven items to assess participants' experience with engaging in transparent research practices. These items were preceded by the statement: "We would like to know about your past research experiences. Please check all that apply." Example (yes/no) items included registering a research study on a publicly available Website, publishing a research protocol before implementing a study, uploading a data set to a publicly available repository, and sharing data sets with colleagues.

#### 3.4.2. General data-sharing concerns

The pretest portion of the online survey also included eight items relating to various concerns researchers might have about sharing their data. Participants read the following introductory statement: "Imagine you were asked to share a research data set that you collected or helped to collect by a researcher you do not know. Using the rating scale below (zero being "not at all concerned," and nine being "extremely concerned"), please tell us how concerned you are about... Using their cursor, participants dragged a button along an unmarked menu bar to indicate an approximate value." Example items covered concerns about whether the data set is ready to share with others, whether the data will be kept securely, whether they would receive recognition for providing the data, how much time would be required to obtain the data, and how the data would be interpreted.

#### 3.4.3. Willingness to share data

Participants were also asked questions about their willingness to share their data, in the context of being asked to participate in a meta-analysis. This section was considered the primary outcome and participants answered these questions after receiving their conditions' data request document. Participants read the following statement before completing this section: "Now that you have read the document, please answer the questions below reflecting on the researcher's request to share the data set associated with the study we referenced in our original email." Respondents were then asked to rate, on a Likert-type scale (moving their cursor along an unmarked menu bar, as before, to indicate an approximate value), their level of agreement with six statements. Example statements included "I would

be interested in sharing my data set for inclusion in this meta-analysis" and "I would be willing to share my data set as long as I am a coauthor of this meta-analysis." The end response options were "do not agree," and "completely agree," and the middle response option was "neither agree nor disagree."

#### 3.4.4. Meta-analysis data sharing concerns

Finally, participants were asked specific questions about their data-sharing concerns with regard to participating in a meta-analysis. Several questions corresponded to concerns explored in the pretest portion of the survey (e.g., how the data will be stored and who will have access). Other questions related to how the data set would be analyzed, whether the findings would be shared with them before publication, how the data would be used, and whether the data set would be reanalyzed later. In addition, respondents rated concerns about being compensated for providing the data set and being included as an advisor on the project. Again, all items were rated using a Likert-type, unmarked, semantic differential scale, with response options "not at all concerned" and "extremely concerned" offered at the end points of the scale and "neither concerned nor unconcerned" offered at the midpoint of the scale.

### 3.5. Data analysis

#### 3.5.1. Baseline equivalence and missing data analysis

We first tested the baseline equivalence of study groups by running bivariate statistical tests, *t*-tests for continuous variables, and chi-squared tests for dichotomous variables. Demographic variables, as well as items related to transparent research practices, were assessed for group equivalence.

To ensure that participants with missing data did not bias the results or imputation procedures, we tested for differences between participants with complete cases and participants with at least one piece of missing information. Participants were required to complete every demographic survey item; therefore, the only missing information occurred in the substantive items. We conducted missing data multiple imputation in R using the package *mi* [16]. To avoid biasing the imputed data, we included all demographic variables in the imputation model and included a condition indicator. The "rhat" convergence diagnostic, which assessed the variance within each chain and then compared it across all chains, indicated a well-fitting model. We exported 10 unique data sets that were used to conduct all further analyses.

#### 3.5.2. Exploratory factor analysis

After data cleaning and imputation procedures, we conducted an exploratory factor analysis on the main pretest measure and outcome measure—participants' attitudes toward data sharing—to assess the factor structure and loadings. We used the *psych* package in R [17] to conduct all factor analyses. To determine the number of factors within

each scale, we used the Very Simple Structure [18] and parallel analyses [19]. Based on the number of factors the models suggested, we next ran an exploratory factor analysis. We constrained the model to the given number of factors determined by the previous analyses and used a varimax rotation to provide greater clarity to the factor loadings. We used the conventional 0.40 as a marker for significant item factor loadings, assuming that the item did not load onto any additional factors above 0.10. We summed the items within each factor to create factor scores.

### 3.5.3. Impact analyses on primary outcome

We then conducted analyses to assess the impact of the “intervention” (i.e., the group that received the hypothetical data-sharing agreement) relative to the “control” (i.e., the group that did not receive the data-sharing agreement). To control for preintervention differences and reduce the impact of potential confounding factors, we used multiple regression models that controlled for: a) participant demographic information (e.g., gender, race, age, and institution), b) participants’ “pretest” scores on their previously held data-sharing concerns and research experiences, and c) the date the original article was published. Group assignment was the main independent variable, and the primary outcome of interest was the factor score related to data-sharing attitudes. We explored model fit and possible confounding variables by including interactions between the intervention variable and all other covariates. We calculated an effect size by dividing the intervention’s regression coefficient by the unadjusted pooled standard deviation [20].

### 3.5.4. Exploratory analyses on secondary outcome

We also analyzed the secondary outcome, focused on the specific concerns about sharing data in the context of a meta-analysis (meta-analysis data-sharing concerns). We did not hypothesize that the 17-item survey fit a traditional factor analytic structure and therefore we analyzed each item individually. To understand how each item affected participants’ data-sharing attitudes, we rank-ordered the control group’s responses, where one represented the largest concern and 17 represented the smallest concern. To assess the data-sharing agreement’s impact on these concerns, we calculated standardized mean-difference effect sizes for each item, rank-ordering each effect size from 1 to 17.

All analyses, unless otherwise noted, were conducted using base R [21] and the R package dplyr [22].

## 4. Results

### 4.1. Group equivalence

A total of 247 individuals consented to participate in the survey (response rate = 20.5%). Of these, 123 had been assigned to the intervention condition and 124 to the control condition. Analyses of baseline differences suggested that

the randomization process successfully produced statistically equivalent groups. As presented in Table 1, we tested for and found no statistically significant differences between the intervention and control groups on demographic characteristics such as age ( $\chi^2 = 2.25$ ,  $P = 0.81$ ), gender ( $\chi^2 = 1.11$ ,  $P = 0.57$ ), race/ethnicity ( $\chi^2 = 0.01$ ,  $P = 0.99$ ), and level of education ( $\chi^2 = 0.01$ ,  $P = 0.99$ ). The groups had similar levels of experience engaging in seven different research practices ( $\chi^2 = 11.03$ ,  $P = 0.20$ ).

### 4.2. Missing data analyses

A total of 51 percent ( $n = 126$ ) of respondents had missing data on at least one item (Online Appendix Table B.1). The proportion of respondents missing data in each group was not statistically significant ( $\chi^2 = 3.40$ ,  $P = 0.06$ ), although a greater proportion of intervention participants (56%) had at least one item missing compared with control participants (44%). The average participant was missing 5.62 items ( $SD = 9.47$ , Median = 1), and the difference between the two groups was also not statistically significant (Mean: IN = 5.59, CO = 5.65;  $t = 0.06$ ,  $P = 0.95$ ).

To ensure that participants with missing data did not bias the results, we assessed the demographic differences between participants who had complete data versus those who had at least one item missing. The results again revealed no significant differences between the two groups ( $P$ s > 0.05). Although no differences were found, we continued to use the multiply imputed data sets for all analyses.

### 4.3. Transparent research practices

As presented in Table 2, before implementing the data-sharing agreement intervention, we asked participants about their past research practices. No significant differences were found between the intervention and control groups ( $P$ s > 0.05). The most commonly reported experience was “sharing a data set with a colleague” (75%). The second most commonly reported experience was to “provide additional study information to a researcher for a systematic review or meta-analysis,” where nearly 72 percent of participants indicated they had provided some information. Only 29 percent of participants, however, had ever provided a data set to a researcher for inclusion in a meta-analysis. Moreover, only 18 percent of participants indicated that their data set had been uploaded to a publicly available repository or Website.

To explore further the transparent practices of researchers, we created two additional aggregate variables. The first variable, “Experience registering a study,” provided the percentage of individuals who answered “yes” to at least one of two items about trial registration (Transparent Research Practices items 1 or 2). No significant differences were found between the two groups and a little more than half of the participants said they had participated

**Table 1.** Characteristics of participants at baseline

Variable	Intervention	Control	Group difference
<i>N</i>	123	124	
Gender			$\chi^2 = 1.11, P = 0.57$
Female	71 (58)	69 (56)	
Male	52 (42)	55 (44)	
Age			$\chi^2 = 2.25, P = 0.81$
25–34	12 (10)	11 (9)	
35–44	31 (25)	29 (24)	
45–55	30 (25)	38 (31)	
56–64	29 (24)	31 (25)	
65–74	18 (15)	13 (11)	
75+	2 (2)	1 (1)	
Race			$\chi^2 = 0.01, P = 0.99$
Nonwhite	12 (10)	12 (10)	
White	106 (90)	104 (90)	
Degree			$\chi^2 = 0.01, P = 0.99$
Ph.D. or equivalent	112 (89)	114 (92)	
Other	14 (11)	10 (8)	
Institution			$\chi^2 = 2.02, P = 0.73$
Primarily Research University	62 (50)	67 (54)	
Primarily Teaching University	22 (18)	24 (19)	
Medical institution	18 (15)	17 (14)	
Research firm	10 (8)	5 (4)	
Other	12 (10)	11 (9)	
Primary area of research			$\chi^2 = 11.03, P = 0.20$
Education	9 (7)	12 (10)	
Criminology	6 (5)	5 (4)	
Medicine	1 (1)	9 (7)	
Psychology	52 (42)	47 (38)	
Public health	41 (33)	32 (26)	
Social work	3 (2)	1 (1)	
Sociology	1 (1)	1 (1)	
Statistics/Methodology	5 (4)	10 (8)	
Other	6 (5)	7 (6)	
Date of study publication	2009.08	2009.49	$t = 1.03, P = 0.31$

Abbreviations: *N*, 247; Group differences are between intervention and control groups; IN, Intervention condition; CO, Control condition; Numbers in parentheses represent percentage.

in one of these two activities (56.28%). The second variable, “Experience providing a data set in some way,” provided the percentage of individuals who answered “yes” to at least one of four items about sharing a data set (Transparent Research Practices items 3, 5, 6, or 7). Again, the two conditions did not differ and more than three-fourths of the participants said they had provided a data set in some capacity (79.76%).

#### 4.4. Factor analysis and primary outcome analysis

The results of the factor analyses on the pretest variable indicated the presence of two primary factors. The results of the factor analysis, and the significant factor loadings,

are presented in [Online Appendix Table B.2](#). Reliability analysis results indicated two well-defined factors ( $\alpha$ s = 0.72 and 0.64). We summed the items’ scale scores to create two factor scores.

We also conducted the same analysis on the primary outcome variable. The results of the Very Simple Structure and parallel analyses again indicated the presence of two primary factors, each reasonably well fitting and of sufficient reliability. The results of the factor analysis, as well as the significant factor loadings, are presented in [Table 3](#). We again summed the item scores to create factor scores for each factor.

Summary statistics of the primary outcome variable are delineated in [Table 3](#). Based on this analysis, the

**Table 2.** Research experiences of study participants at baseline

I have...	Overall (%)	IN (%)	CO (%)	Group difference
(Q1) Registered a research study on a site such as <i>clinicaltrials.gov</i> or another publicly available Website.	29.84	29.03	30.65	$\chi^2 = 0.02, P = 0.89$
(Q2) Published a research protocol/plan in a journal or publicly available Website before implementing a study.	52.02	50.00	54.03	$\chi^2 = 0.26, P = 0.61$
(Q3) Uploaded a research data set to a publicly available repository.	18.15	14.52	21.77	$\chi^2 = 1.74, P = 0.19$
(Q4) Provided additional study information to a researcher for a systematic review or meta-analysis.	71.78	67.74	75.81	$\chi^2 = 1.61, P = 0.20$
(Q5) Provided a research data set to be included in a meta-analysis.	29.03	24.19	33.87	$\chi^2 = 2.37, P = 0.12$
(Q6) Shared a research data set with a colleague.	75.00	77.42	72.58	$\chi^2 = 0.54, P = 0.46$
(Q7) Shared a research data set with a researcher who I do not know.	21.78	17.74	25.81	$\chi^2 = 1.92, P = 0.17$
(Q8) Experience registering a study	56.28	54.47	58.06	$\chi^2 = 0.19, P = 0.66$
(Q9) Experience providing a data set in some way (i.e., yes to questions 3, 5, 6, or 7).	79.76	82.11	77.42	$\chi^2 = 0.58, P = 0.45$

Abbreviations: N, 247; Group differences are between intervention and control groups; IN, Intervention condition ( $n = 123$ ); CO, Control condition ( $n = 124$ ).

intervention produced a statistically significant impact on the participants' attitudes toward sharing their data set, and the magnitude of this effect was fairly large ( $d = 0.77, 95\% \text{ CI } [0.51, 1.03]$ ). We used a multiple regression model to assess the impact of the intervention while controlling for additional variables (Table 4). The results revealed a statistically significant positive effect in favor of the intervention group ( $b = 4.31, \text{ SE} = 0.75, t = 5.73, P < 0.001$ ). Pretest adjusted effect size calculations revealed a large, statistically significant effect size ( $d = 0.65, 95\% \text{ CI } [0.39, 0.90]$ ). Exploratory models that included interaction terms indicated that the original model fit the data well (Online Appendix B.3–B.5). Therefore, we considered the original model to be the final impact analysis model.

Following our research protocol, we explored the potential moderation of intervention effects based on participant characteristics by conducting six subgroup analyses. The results revealed that the intervention effect varied across several subgroups, but the differences were not statistically significant. The results are presented in Online Appendix B.6.

#### 4.5. Exploratory analysis of secondary outcome

In addition to asking participants about their willingness to supply the primary study data set for the meta-analysis, we also asked participants about specific data-sharing concerns in the context of being asked to share their data for a meta-analysis (Supplemental Table B.7). To understand how the data-sharing agreement intervention affected the

individual concerns, we calculated standardized mean-difference effect sizes for each item. We then rank-ordered the effect sizes and compared the rankings with the rank-order of control group concerns. Looking specifically at the top six concerns, the data-sharing agreement had an intervention impact on three of the six concerns. The largest effect size ( $d = -0.45, 95\% \text{ CI } [-0.69, -0.21]$ ) was found for the second-ranked concern (“How will the data be used after completion of the meta-analysis?”).

## 5. Discussion

The purpose of this study was to determine whether and how much a data-sharing agreement, sent to primary study authors when requesting a primary study data set, affected primary study authors' attitudes toward their willingness to share IPD. A total of 247 individuals provided informed consent and completed at least the demographic questionnaire. We asked participants about their past research experiences, general concerns about data sharing, their willingness to share their data set, and the concerns preventing them from sharing their data set.

The results indicated a positive and statistically significant intervention effect after controlling for potential confounding factors ( $d = 0.65, 95\% \text{ CI } [0.39, 0.90]$ ). Translated to the What Works Clearinghouse's Improvement Index [23], this result indicated that should control group members receive the intervention, we would expect

**Table 3.** Exploratory factor analysis of primary outcome

Factor	Variable	Loading		Reliability	Descriptive statistics		
		F1	F2		In	CO	d
Sharing my data set <sup>b</sup> .	I would be interested in sharing my data set for inclusion in this meta-analysis.	0.83		0.79	16.99 (5.88)	12.18 (6.57)	0.77 (0.51, 1.03)
	I would not be willing to share my data set for inclusion in this meta-analysis. <sup>a</sup>		0.76				
	Now that an e-mail has been sent to me asking for the data set, I would share my data set for inclusion in this meta-analysis.	0.66					
Sharing my data set with conditions.	I would be willing to share my data set as long as I am a coauthor of the meta-analysis.		0.99	0.47	4.26 (3.91)	4.21 (3.85)	0.01 (−0.25, 0.27)
	I would be willing to share my data set as long as I receive financial compensation.		0.46				

Abbreviations: N, 247; IN, intervention condition ( $n = 123$ ); CO, control condition ( $n = 124$ ).

<sup>a</sup> Item reverse-scored before calculations.

<sup>b</sup> Represents the primary outcome; Cronbach's alpha reported for factor 1; bivariate correlation reported for factor 2.

that their willingness to share the primary data set would improve 24 percent. In other words, a member of the control group is 24 percent more likely to share her data set should she receive the data-sharing agreement.

### 5.1. Changes to the original data-sharing agreement

The results of this study provide evidence to support the hypothesis that a data-sharing agreement makes an impact on primary study authors' attitudes toward their willingness to share IPD. At a minimum, therefore, we suggest that meta-analysts interested in using IPD in their analyses send our data-sharing agreement or something similar. What is also clear, however, is that concerns about sharing IPD

among primary study authors remain. We believe that certain changes to the data-sharing agreement will reduce these concerns, and we have therefore made several changes of note (see Online Appendix A).

The primary concern about sharing the data set was the storage and accessibility of the data set, based off the pretest factor (“Concerns with data set accessibility”) and the secondary outcome concerns. Thus, we clarified language from the previous version and moved this agreed-upon statement to the top of the list. A secondary concern with data set sharing was the future use of the data set, primarily whether the data set will be used to conduct additional analyses beyond those specified for the meta-analysis.

**Table 4.** Impact analysis of primary outcome

Variable	Coding	Coefficient (SE)	P-value
Intercept	—	14.35 (2.65)	0.001
Intervention	0 = CO, 1 = IN	4.31 (0.75)	0.001
Gender	0 = female, 1 = male	−0.09 (0.79)	0.914
Race	0 = white, 1 = nonwhite	0.49 (1.24)	0.691
Age	Continuous variable	0.14 (0.35)	0.690
Degree type	0 = non-Ph.D., 1 = Ph.D.	−1.32 (1.35)	0.328
Date of original study publication	Continuous variable	0.02 (0.14)	0.869
Experience providing data in some way	0 = no, 1 = yes	3.90 (0.95)	0.001
Concerns about the condition of the data set (pretest factor 1)	Continuous variable	0.08 (0.06)	0.227
Concerns about data set accessibility (pretest factor 2)	Continuous variable	−0.43 (0.10)	0.001

Abbreviations: N, 247; IN, intervention condition ( $n = 123$ ); CO, control condition ( $n = 124$ ); SE, standard error; Intervention effect size calculated by dividing the regression coefficient by the pooled standard deviation ( $d = 0.65$ ).

We also added additional statements to address specific concerns articulated by the results. For example, a primary concern of the participants was whether there were errors in the data set or whether the codebook was available, in addition to the fourth-ranked concern (“The time it takes to prepare the data set”). To address these concerns directly, we added a statement about the “Dataset and Codebook Completeness,” which says: “The Primary Researcher (you) is under no obligation to modify, edit, update, or re-analyze the primary study dataset. Please provide, however, any documentation available that may be useful to the meta-analysis.” This ensures that the primary study authors feel comfortable with the process and not overburdened. To ensure that accurate information is collected, however, we added the second statement asking primary study authors to send any additional helpful information. We anticipate that this will address their concerns.

Finally, we added three additional statements to address specific smaller concerns with data set sharing. The “Right to Consultation” statement says: “The Primary Researcher (you) has the right to provide input and consultation to the project.” We believe this statement will provide further assurances to the primary study authors about their participation in the project. The “Right to Collaboration” statement addresses coauthorship, should significant participation warrant such involvement. Finally, although the results indicated financial compensation was not a significant concern, we added a statement that addresses financial reimbursement directly.

### 5.2. *Implications for practice*

It is clear that meta-analysts who seek IPD from primary study authors using email communication should send a data-sharing agreement. Should meta-analysts choose not to use our version, at a minimum, their data-sharing agreement should address the concerns indicated by our results. These include how the data will be stored, how it will be used for the current analysis, and in the future, who will have access to the data set, and whether the primary study author will be involved in the project, either as an adviser or coauthor.

We also believe that meta-analysts should consider requesting the IPD from primary study authors for each primary study included in the meta-analysis, regardless of whether an IPD meta-analysis is planned. Although we recognize this shift has the potential to create logistical challenges, we make this suggestion because of the wealth of benefits available. The primary issue of requesting the data sets regards the time it takes simply to find the primary study author and their e-mail for each included study. Thanks to powerful Internet search engines and professional social media Websites (e.g., LinkedIn, ResearchGate), it is easier than ever before to locate a primary study author. Moreover, in our experience, most meta-analyses require inquiry from primary study authors for traditionally missing information, for example, the percentage of males in the intervention group or the number of

participants in each condition. Emailing each primary study author, therefore, may not increase the required time already necessary to conduct a high-quality review.

### 5.3. *Limitations*

Although the results of this study are encouraging, they are not without considerable limitations. Primarily, this study relied on the self-reported attitudes of primary study authors and not their behaviors. We believe that the scales we constructed represent these behaviors, in as much sense as any attitudinal scale can approximate behaviors. Of course, it is quite different to say you are willing to share your data set than to do it. It is not lost on us that this study did not address this primary concern.

In addition, our results should not be generalized to data-sharing attitudes or practices in general. The purpose of this project, stated clearly throughout the survey, was to test the effectiveness of the data-sharing agreement in the context of sharing data for use in a meta-analysis. Researchers who wish to collect IPD for the purposes of reanalysis may be met with significant and different challenges relative to meta-analysts interested in the data set for use in a meta-analysis. Future researchers should consider adapting our data-sharing agreement for use in these circumstances.

## 6. **Conclusions**

We believe that sharing data among researchers is not only an ethically sound choice but also one that has the potential to make significant improvements in future research. The implications of improving research have profound impacts on the individuals that researchers and meta-analysts seek to serve. As such, the use of data-sharing agreements should be considered a priority among meta-analyst and should be a default application in the future.

## **Acknowledgment**

This project was supported by Award No. 20153030 and 20161821, awarded by the University of California-Berkeley, Berkeley Initiative for Transparency in Social Sciences (BITSS). The authors thank Drs. Mark Lipsey, Terri Pigott, and Jessaca Spybrook for their help in reviewing the project materials.

## **Supplementary data**

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jclinepi.2018.10.006>.

## **References**

- [1] Shamoo AE, Resnik DB. *Responsible conduct of research*. New York, NY: Oxford University Press; 2009.

- [2] Polanin JR, Williams RT. Overcoming obstacles in obtaining individual participant data for meta-analysis. *Res Synthesis Methods* 2016;7:333–41.
- [3] Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *JAMA* 2015;313:793–4.
- [4] Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes* 2012;5:238–40.
- [5] Jaspers GJ, Degraeuwe PLJ. A failed attempt to conduct an individual patient data meta-analysis. *Epidemiology* 2008;19:640–8.
- [6] Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in the *BMJ* and *PLOS Medicine*. *BMJ* 2018;360:k400.
- [7] Polanin JR. Efforts to retrieve individual participant datasets for use in a meta-analysis result in moderate data sharing but many datasets remain missing. *J Clin Epidemiol* 2018;98:157–9.
- [8] Villain B, Dechartres A, Boyer P, Ravaud P. Feasibility of individual patient data meta-analyses in orthopaedic surgery. *BMC Med* 2015;13:1–10.
- [9] Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ* 2012;344:d7762.
- [10] Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijls L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008;27:1870–93.
- [11] 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.
- [12] Pigott TD, Williams RT, Polanin JR. Individual participant data meta-analysis with correlations. *Res Synth Methods* 2012;3:257–68.
- [13] Tudur-Smith C, Hopkins C, Sydes MR, Woolfall K, Clarke M, Murray G, et al. How should individual participant data (IPD) from publicly funded clinical trials be shared? *BMC Med* 2015;13:298–306.
- [14] Campbell EG, Clarridge BR, Gokhale M, Birenbaum L, Hilgartner S, Holtzman NA, et al. Data withholding in academic genetics: evidence from a national survey. *JAMA* 2002;287:473–80.
- [15] Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ* 2017;357:J1390.
- [16] Gelman A, Hill J, Su Y, Yajima M, Pittau M, Goodrich B, et al. Missing Data Imputation and Model Checking (Version 1.0) [Software] 2015. Available at <https://cran.r-project.org/web/packages/mi/mi.pdf>. Accessed June 1, 2017.
- [17] Revelle W. Procedures for Psychological, Psychometric, and Personality Research (Version 1.7.3.21) [Software] 2017. Available at <https://cran.r-project.org/web/packages/psych/psych.pdf>. Accessed June 1, 2017.
- [18] Revelle W, Rocklin T. Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behav Res* 1979;14(4):403–14.
- [19] Velicer W. Determining the number of components from the matrix of partial correlations. *Psychometrika* 1976;41(3):321–7.
- [20] Polanin JR, Snijlsteit B. Campbell methods policy note on converting between effect sizes (Version 1.1, updated December 2016). Oslo: The Campbell Collaboration; 2016.
- [21] R Core Team. R: A language and environment for statistical computing (Version 3.4.2) [Software] 2018. Available at <http://www.R-project.org/>. Accessed June 1, 2017.
- [22] Wickham H, Francois R. A Grammar of Data Manipulation (Version 0.5.0) [Software] 2016. Available at <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>. Accessed June 1, 2017.
- [23] U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. What Works Clearinghouse Procedures and Standards Handbook (Version 3.0) 2014. Available at <http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19>. Accessed June 1, 2017.