



## Intent to share Annals of Internal Medicine's trial data was not associated with data re-use

Claude Pellen, Laura Caquelin, Alexia Jouvance-Le Bail, Jeanne Gaba,  
Mathilde Vérin, David Moher, John P A Ioannidis, Florian Naudet

### ► To cite this version:

Claude Pellen, Laura Caquelin, Alexia Jouvance-Le Bail, Jeanne Gaba, Mathilde Vérin, et al.. Intent to share Annals of Internal Medicine's trial data was not associated with data re-use. Journal of Clinical Epidemiology, 2021, 137, pp.241-249. 10.1016/j.jclinepi.2021.04.011 . hal-03222940

**HAL Id: hal-03222940**

**<https://hal.science/hal-03222940>**

Submitted on 14 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Highlights / What is new:**

- Published re-uses are mainly secondary analyses. Only a few meta-analyses of individual participant data are performed, and no re-analyses were identified.
- Although more than half of the articles mention willingness to share their datasets, no significant association was found between intention-to-share and published re-uses of the data.
- The ICMJE policy needs to implement an evaluation component in its new policy.

**INTENT TO SHARE ANNALS OF INTERNAL MEDICINE'S TRIAL DATA WAS  
NOT ASSOCIATED WITH DATA RE-USE**

**Claude Pellen, MA<sup>1</sup>, Laura Caquelin, MA<sup>1</sup>, Alexia Jouvance-Le Bail, MA<sup>1</sup>, Jeanne Gaba, MA<sup>1</sup>,  
Mathilde Vérin, MA<sup>1</sup>, David Moher, PhD<sup>2</sup>, John P. A. Ioannidis, MD, DSc<sup>3,4,5</sup>, Florian Naudet,  
MD, PhD<sup>1</sup>**

<sup>1</sup>Univ Rennes, CHU Rennes, Inserm, CIC 1414 [(Centre d'Investigation Clinique de Rennes)], F-35000 Rennes, France

<sup>2</sup>Center for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

<sup>3</sup>Department of Medicine, Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, United States

<sup>4</sup>Departments of Epidemiology and Population Health and of Biomedical Data Science, Stanford University School of Medicine, Stanford, United States

<sup>5</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, United States

**Corresponding author:** Claude Pellen ([cld.pellen@gmail.com](mailto:cld.pellen@gmail.com)); +33 2 99 28 43 21; CIC 1414, 2 rue Henri le Guilloux, 35 000 Rennes, France

**ABSTRACT**

**Objective:** To explore the impact of the Annals of Internal Medicine (AIM) data-sharing policy for randomized controlled trials (RCTs) in terms of output from data-sharing (i.e. publications re-using the data).

**Study design and setting:** Retrospective study. RCTs published in the AIM between 2007 and 2017 were retrieved on PubMed. Publications where the data had been re-used were identified on Web of Science. Searches were performed by two independent reviewers. The primary outcome was any published re-use of the data (re-analysis, secondary analysis, or meta-analysis of individual participant data [MIPD]), where the first, last and corresponding authors were not among the authors of the RCT. Analyses used Cox (primary analysis) models adjusting for RCTs characteristics (registration: <https://osf.io/8pj5e/>).

**Results:** 185 RCTs were identified. 106 (57%) mentioned willingness to share data and 79 (43%) did not. 208 secondary analyses, 67 MIPD and no re-analyses were identified. No significant association was found between intent to share and re-use where the first, last and corresponding authors were not among the authors of the primary RCT (adjusted hazard ratio = 1.04 [0.47-2.30]).

**Conclusion:** Over ten years, RCTs published in AIM expressing an intention to share data were not associated with more extensive re-use of the data.

**Keywords:** Data sharing, Open science, Clinical Research, Randomized Controlled Trial, Re-use of data, Meta-research

**Running title:** Intent to share trial data and data re-use

**Word Count:** 3985

**Title:** Intent to share Annals of Internal Medicine's trial data was not associated with data re-use

## 1. INTRODUCTION

Data-sharing (i.e. sharing of data, codes, programs and material) is the norm in many scientific disciplines, but until recently this has not been the case with biomedical research (1). In medicine, randomized controlled trials (RCTs) are considered influential and therefore availability of their data is expected to be crucial in the evaluation of health interventions (e.g. for meta-analyses of individual participant data, MIPD). In June 2017, the International Committee of Medical Journal Editors (ICMJE) required a data-sharing plan to be included in each paper (and pre-specified in study registration) (2). As these new requirements for publishing experimental findings come into effect, it is necessary to assess whether they have their intended effects.

Among the leading general medical journals, the New England Journal of Medicine (NEJM), The Lancet, the JAMA and JAMA Internal Medicine have had no specific policy for sharing data from RCTs until recently. The British Medical Journal (BMJ) and the Public Library of Science (PLOS) Medicine have already adopted stronger policies, reaching beyond the ICMJE policy, which render data-sharing mandatory with the publication of RCTs. Nevertheless, in view of actual data-sharing rates their policy seems imperfect (3).

The Annals of Internal Medicine (AIM) has encouraged (but not required) data-sharing since 2007 (4). Since then, the journal has required a reproducible research statement to be included in every original research article (including RCTs). This reproducible research statement indicates “whether the study protocol, data, or statistical code is available to readers, and under what terms authors will share this information.” While this policy did not make data-sharing mandatory, its aim was “to help the scientific community evaluate, and build upon, the research findings” published in AIM. Importantly, this is to a large extent what is required by the new ICMJE policy. Therefore, a retrospective analysis of RCTs published in AIM between 2007 and 2017 could provide a proxy for the expected impact of the ICMJE policy.

We explored the effectiveness of RCT data-sharing from AIM publications in terms of output from data-sharing (i.e. publications where the data has been re-used). We specifically aimed to describe the data-sharing practices in RCTs published in AIM over a decade (2007-2017), to assess the association between intent to share and published re-uses of data, and to assess the association of intent to share with citation rates.

## 2. METHODS

The methods were specified in advance. They were documented in a protocol registered with the Open Science Framework (OSF) on 3rd August 2018 (<https://osf.io/gnt6u/>).

## **2.1. ELIGIBILITY CRITERIA**

We surveyed a retrospective cohort of RCTs published in the AIM between April 2007 and December 2017. The RCTs included the following designs: two parallel groups and multiple groups, cluster trials and cross-over studies, non-inferiority and superiority trials. All publications were inspected to exclude secondary analyses and re-analyses of a previously published RCT. Publications without reproducible research statements (i.e. those that did not comply with the policy) were excluded.

## **2.2. SEARCH STRATEGY AND SELECTION OF PRIMARY PUBLICATIONS OF STUDIES IN AIM**

We identified eligible studies from PubMed/Medline using the following strategy: (annals of internal medicine) AND ("2007/04/01"[Date - Publication]: "2017/12/31"[Date - Publication]) / limitation randomized controlled trial.

Two reviewers (CP and AJLB) performed the eligibility assessment independently. Disagreements were resolved by consensus or in consultation with a third reviewer (FN).

## **2.3. SEARCH STRATEGY AND STUDY SELECTION FOR PUBLISHED RE-USES**

We used the Clarivate Web of Science database to identify secondary publications derived from these primary trials, since it would be extremely unlikely for a secondary publication not to cite the primary trial. We identified and recorded the total number of citations for each primary publication. We recorded the number of citations by articles that used individual patient data from the primary article in the Annals. For articles considered as potentially secondary publications, the abstracts and, when necessary, full texts were examined by two independent reviewers (among LC, JG, and MV) to confirm eligibility. One reviewer inspected all included citations and when he disagreed with the inclusion, disagreements were resolved in consultation with a third reviewer (FN). In addition, whenever a methodological article was cited in the primary article in the Annals (i.e. an article describing the methods and protocol of a trial), we entered this article in the Web of Science searches to identify additional citations.

## 2.4. DATA EXTRACTION

A data extraction sheet was developed. For each article included, we extracted information on study characteristics (date of publication, country (USA/Europe/Asia/other), intervention type (drug/device/complex intervention), control group (active/inactive), medical specialty (medicine/surgery/psychiatry), total sample size, result on the primary outcome (positive/negative), and funding source (academic/industry/charity/mixed). Detailed information on the data-sharing plan was extracted from the reproducible research statement. We recorded whether a statement indicated that the data was available (i.e. intent to share data). If the authors intended to share data, we extracted the type of data-sharing plan (1/ directly available, 2/ available upon request) and the type of material that was intended to be shared (1/ data-set, 2/ code, 3/ study protocol). We relied on information declared by the authors in the “reproducible research statement”. In case of study protocols, intention to share by authors can include full text protocols and/or protocols available only in registries. If the authors did not intend to share data, we extracted reasons for not sharing data (when available).

Two authors (CP and AJLB) independently extracted the data from the studies included. Disagreements were resolved by consensus or in consultation with a third reviewer (FN).

## 2.5. OUTCOMES

Our primary outcome was the re-use of data (yes/no) documented in all the citing articles on Web of Science. This is a composite outcome defined by any secondary use 1/ in a re-analysis, 2/ in a secondary analysis and 3/ in a MIPD (we used a broad definition including both systematic reviews with MIPD and pooled analyses without systematic review). The primary outcome was limited to published re-uses where the first, last and corresponding authors were not among the authors of the primary article in the Annals.

The pre-specified secondary outcomes were as follows: 1/ Components (1, 2, and 3) of the primary outcome; 2/ Re-use of the data (yes/no) documented in all the citing articles, on Web of Science without the author restriction used for the primary outcome; 3/ Number and type of secondary use (re-analyses, secondary analyses, MIPD) in which the lead authors (first, corresponding and last) of the published re-use of the data were outside the team authoring the primary article in the Annals; 4/ Number and type of secondary use (re-analyses, secondary analyses, MIPD) by an independent team (no author in common); 5/ Number and type of secondary use (re-analyses, secondary analyses, MIPD) by a team where <50% of authors were among the authors of the primary article in the Annals; 6/ Number and type of secondary use (re-analyses, secondary analyses, MIPD) by a team

where  $\geq 50\%$  of authors were among the authors of the primary article in the Annals; 7/ Number and type of secondary use (re-analyses, secondary analyses, MIPD) by a team where 100% of authors were among the authors of the primary article in the Annals; 8/ Mention in the published re-uses of the availability of the data re-used (yes/no) and how it was used (qualitative data); 9/ Number of citations (on Web of Science).

## 2.6. STATISTICAL ANALYSIS

A detailed statistical analysis plan was registered with the OSF on 1<sup>st</sup> August 2019 (<https://osf.io/zck9y/>) prior to merging the AIM primary articles database and the re-use database. The general strategy for modeling was as follows: we first used a univariate model with fixed effects to explore solely the relationship between the dependent and the explanatory variables, adjusting on possible confounders (country, intervention type, control group, medical specialty, result (positive/negative) on the primary outcome, funding source, and sample size). All associations with a p-value  $< 0.25$  were subsequently explored in a multivariate model. In all final models, we considered adding a random effect to account for studies that were published by the same team (defined as groups of primary AIM articles clustered by any common authors).

We described ‘intent to share data’ rates over time using a logistic regression (model diagnostics checked for linearity, multicollinearity and absence of influential observations). Results are expressed as odds ratios (OR) with their 95% confidence intervals (CIs). The primary outcome and all its components were analyzed using Kaplan-Meier test and compared using proportional hazards regression (model diagnostics checked for linearity, proportional hazards and absence of influential observations). Results are expressed as hazard ratios (HR) with their 95% CI. Other secondary outcomes were analyzed as count data using Poisson or (when necessary) binomial negative regressions (model diagnostics checked for linearity and absence of overdispersion). Results are expressed as incidence rate ratios (IRR) with their 95% CI. Model diagnostics are detailed in the statistical analysis plan and their results can be accessed on the OSF (<https://osf.io/8pj5e/>).

All statistical analyses were performed at a 5% significance level (two-sided tests). All analyses were performed under R version 3.4.1 (2017-06-30, The R Foundation for Statistical Computing).

## 2.7. CHANGES FROM THE INITIAL PROTOCOL

The following clarifications were described prior to the analyses in the statistical analysis plan: 1/ we decided to classify pooled analyses as MIPD because this point was unclear in the initial



protocol, 2/ we made it clear that the year of publication (and date) was used to calculate each follow-up time (date of publication to date of citation searches) in survival models and as an offset in count models, and not as a potential confounder and 3/ also, we excluded citations concerning long-term follow-up and use of stored biological samples, since these analyses obviously implied the acquisition of new data.

We considered the sample size of the primary article in the Annals as a potential confounder, since it could be responsible for changes in both the intent to share and the number of re-uses, and therefore could be an important confounding factor, which we omitted to plan for in our initial protocol.

While we initially planned to include 3 types of re-use (secondary analyses, re-analyses and MIPD), we decided to include meta-analyses of aggregate data as supplementary analyses. Indeed, data availability can have an impact when the initial publication does not report the data required to perform the meta-analysis.

## 2.8. ROLE OF THE FUNDING SOURCE

This work was supported by the Fondation pour la Recherche Médicale, grant number 6616 to Claude Pellen, and Région Bretagne Boost ERC grant (18HC432-01N). Work by FN on data-sharing is supported by a grant from the French National Research Agency – ANR (Reproducibility in Therapeutic Research / ReITheR: ANR-17-CE36-0010-01). The funders were not involved in the study design, data collection, analysis, interpretation, or writing of the manuscript.

## 3. RESULTS

### 3.1. CHARACTERISTICS OF PRIMARY ARTICLES AND PUBLISHED RE-USES

**Figure 1** shows the study selection process. The searches carried out on 27<sup>th</sup> May 2018 retrieved 257 items. Of these, 185 were primary articles and had a data-sharing statement; these articles were included. These RCTs had a median sample size of 299 participants (interquartile range (IQR) [158-719]), they had academic funding in 79 cases (42.7%), and were led by teams from USA in 97 cases (52.4%). Eighty-eight RCTs (47.6%) evaluated drug interventions, 89 (48.1%) complex interventions (e.g. psychotherapeutic program), and 8 (4.3 %) evaluated devices. Based on the data-sharing statements in the articles, 106 studies mentioned willingness to share the dataset and 79 mentioned that the data were not available. Of these, none provided a reason for not sharing. **Table 1** presents the characteristics of these studies displayed by intent to share or not.

Search for re-uses of these primary articles were carried out on 9<sup>th</sup> August 2018. There were 15,378 citations identified through Web of Science searches (14,576 from primary articles and 802 more from protocols of these studies). Among these, we identified 208 secondary analyses, 67 MIPD, and 0 re-analyses. We also identified 407 meta-analyses of aggregate data.

### 3.2. INTENT TO SHARE TRENDS OVER TIME

**Figure 2** shows data-sharing trends over time. We found no association between ‘intent to share data’ rates and time (years) (OR = 1.00 [0.90 -1.11], adjusted OR = 0.98 [0.84-1.13]). Details of the adjusted model are provided in **Web-appendix 1**. Proportions of code and protocol availability were always higher than levels of intent to share data, and reached 100 % for protocol availability in 2017.

### 3.3. ASSOCIATION BETWEEN INTENT TO SHARE AND PUBLISHED RE-USES OF DATA

Over 10 years, 185 trials were available. Among the 106 of them mentioning willingness to share the dataset, 14 (13.2%) had at least one published re-use where the first, last and corresponding authors were not among the authors of the original RCT. Among the 79 others, 12 (15.2%) had at least one published re-use where the first, last and corresponding authors were not among the authors of the original RCT.

Univariate and multivariate analyses (**Figure 3**) identified no significant association between intent to share and the primary outcome, that is to say published re-uses of the data where the first, last and corresponding authors were not among the authors of the primary article in the Annals (adjusted HR = 1.04 [0.47-2.30]), nor was there any association with its different components (adjusted HR = 0.96 [0.32-2.90] for secondary analyses, adjusted HR = 1.23 [0.37-4.06] for MIPD). The same results were observed in analyses using no author restriction criteria (adjusted HR = 1.30 [0.84-2.01]). Supplementary analyses, including meta-analyses of aggregate data not specified in the registered protocol, are presented in **Web-appendix 2**. The results were consistent with those of the main analysis.

**Figure 4** presents count outcomes related to the number of re-uses (by different types of re-use). A few statistically significant associations were found in univariate analyses but they were not confirmed in multivariate analyses for pre-specified outcomes. Only the outcome “MIPD/pooled analyses with no author in common with the primary RCT” retained a weak association signal in multivariate analyses. The number tended to be greater when there was no mention of data-sharing

(adjusted IRR = 0.03 [0.00-0.33]). A careful examination of individual papers found that among the 13 re-uses of RCTs stating they have no intention to share, 12 had the same funder as the primary RCTs, all of these being sponsored by pharmaceutical firms (**Web-appendix 3**). These results are presented in detail in **Web-appendix 4**, and the number of re-uses per RCT are presented in **Web-appendix 5**.

### 3.4. ASSOCIATION BETWEEN INTENT TO SHARE AND NUMBER OF CITATIONS

Univariate and multivariate analyses identified no significant association between intent to share and the number of citations, with 134 citations per 10 publication-years for papers with an intention to share statement and 149 per 10 publication-years for papers stating they had no intention to share (IRR = 0.83 [0.67-1.02]; adjusted IRR = 0.90 [0.72-1.13]).

## 4. DISCUSSION

### 4.1. STATEMENT OF PRIMARY FINDINGS

Over ten years, one might have expected a progressive increase in data-sharing practices with a progressive improvement in data-sharing intent over time. However, and in line with a previous exploration of intent to share data for all research articles in AIM between 2008 and 2012 (5), we did not find any such increase. Interestingly, while Laine et al. (6) found that the intention to share protocols was lower than for the intention to share datasets, we found the opposite, with higher, increasing rates of intention to share protocols, reaching 100% in 2017. This result could be explained by our specific focus on clinical trials. Protocol availability is a core feature of transparency for these studies, especially in the context of the strong, justified external pressure by some watchdog groups. For instance, in response to the efforts by the Centre for Evidence-Based Medicine Outcome Monitoring Project (COMPare) (7) to document outcome switching in RCTs, the Annals editor recalled that the journal routinely asks authors of clinical trials to submit their protocols with their manuscripts, and examines trial registries for the initial and final information entered about trials (8). In data-sharing statements, intention to share codes was always more prominent than the sharing of data. This result is curious, since codes without the data are most often useless.

While more than 55% of the primary articles surveyed intended to make their data available, no association was found between intent to share and re-use of the data. Interestingly, there were only a few published re-uses of the primary articles, and three-quarters were secondary analyses. MIPD were the remaining re-uses and we found no re-analyses, meaning that none of the re-uses of the data was for reproducibility purposes. This last result is unfortunate given the increased deployment of reproducibility checks in other fields (9). However, this observation is in line with an earlier survey

(10) exploring requests for data access from the National Heart, Lung and Blood Institute data repository. Over 16 years, 100 trials were available, 88 of them had a data request and 47 had at least one publication resulting from data-sharing. More than 80% of requests were for secondary analyses or methodological developments, and only 7% were for MIPD. In this survey, only two requests concerned re-analyses. And indeed, there is almost no culture of performing and publishing re-analyses in the clinical trial literature (11), especially by entirely independent authors. Perhaps fear of “hostile” re-analyses or even the possibility that any re-analyses are liable to obtain different results could explain lack of data sharing for this specific purpose. Another survey of cardiometabolic clinical trials available on the clinical study data request platform (12) suggested that despite efforts to make data available, re-uses were rare. Over 4 years, only 3 re-uses had been published among the 537 studies available for access. Similarly, over 11 years (13), only 14 re-uses (5 secondary analyses and 9 MIPD) were published using 51 clinical trial datasets available on the Data Share platform of the National Institute on Drug Abuse. Altogether, these results suggest that, despite some important efforts, data-sharing policies does not systematically achieve large numbers of published outputs.

Different factors could explain this finding. First, the available datasets may not be requested (and used) by external researchers. Second, primary authors may refuse to share their data as promised in their data-sharing statement. For instance, in our survey, only one publication had its dataset directly available on a repository, while all the others were available upon request. Similar rates were observed in a survey (6) of the Annals’ first year of the reproducible research statement policy, and in a second survey (5) accounting for articles from 2008 to 2012. When data is available upon request, data retrieval could be suboptimal as suggested by a previous survey of studies published in PLOS Medicine or BMJ under a strict data-sharing policy (3). However, among 90 authors of trials published between January 1, 2012 and March 1, 2016 in PLOS Medicine, The BMJ and the Annals (14), half of the respondents had a data-sharing plan ( $n = 49$ ) and about one third reported they had received at least one data-sharing request, and very few of these were reported as being refused. It is therefore possible that data is in fact shared, but that re-use of this data will not translate into published output. A cross-sectional web-based survey (15) of the NIH central database repository found that only 67% of the re-uses were published. In addition, shared datasets could serve for pedagogical purpose or for designing trials (e.g. sample size calculation) or any other activity that may not necessarily lead to any published output. However, it is also possible that data-sharing enables numerous secondary analyses to be run, among which only a few reach published posterity. This last hypothesis is of concern, because sub-optimal reporting and selective publication of these re-uses could lead to non-reproducible research (16). In future research, qualitative interviews, including interviews with academic researchers, industry researchers and re-users, could add a finer-grained understanding of these issues.

## 4.2. LIMITATIONS

Our results should be interpreted cautiously. The observational design of the study does not enable us to draw causal inferences between intent to share and research outputs from actual re-use of data. Confounding is a major issue in observational research, and there is no perfect way to handle it. Despite the adjustment of the analyses to various available factors that might influence the patterns of re-use, some factors could not be fully accounted for. For instance, we were not able to adjust for subtle variations that can arise across different medical subspecialties or topics. Of course, “hot and/or controversial topics” (e.g. vaccines and autism, saturated fat and heart disease) can be associated with both intent to share and re-use, but there is no consensual way to measure this parameter. Also, we did not identify the corresponding author’s gender or career stage, while a junior investigator could be more aware of open science practices such as data-sharing. Therefore, this study is prone to residual confounding. In our study, the identification of published re-uses proved to be a difficult task, with the risk of missing some studies. We minimized this risk by using a dual, independent extraction.

Also, it was difficult to define independent re-uses objectively. For instance, the primary article authors could require to be among authors of the re-use as part of their data-sharing statement, even if they were not actively involved in the re-use. This could have introduced some classification bias in our study. For instance, the only signal we identified was for IPD meta-analyses (or pooled analyses) with no author in common with the primary RCT, which were more frequently RCTs without intention to share data. The topics of these RCTs were long-term efficacy of dapagliflozin for patients with diabetes mellitus, adalimumab in moderate to severe Hidradenitis suppurativa, tofacitinib in associations for patients with active rheumatoid arthritis, varenicline in smoking cessation, and the Grazoprevir-Elbasvir combination in Chronic Hepatitis C Virus. The RCT that shared their data was an equivalence trial among treatment-naïve volunteers infected with HIV-1. All but one of these re-uses were sponsored by the same funder as the original trial. It is therefore likely that a given funder decided not to release the data and to perform their own research agenda, including series of pooled analyses. Also, classification bias could be differential between studies stating they do or do not intent to share data. For this reason, we used a restriction on the lead authors (first, last and corresponding) for our primary outcome, and performed a series of sensitivity analyses with various definitions. Importantly, all these analyses yielded very similar results.

Most importantly, this retrospective study explores an early experience of data-sharing. The primary articles included in our study cover 10 years during which the AIM pioneered clinical data-sharing, with a few other journals. As a result, re-uses like MIPD needing more than only one study dataset could have been hard to perform, because the overall environment was unfavorable. Up to

2015 most MIPDs were incomplete, half retrieved less than 80% of the eligible individual participant data and retrieval rates across published MIPDs did not improve through these years (17). Improved access to data of a single study may be of limited interest in this environment. Therefore, the generalization of the results obtained from 2007 to 2017 from a single journal to all journals implementing the new ICMJE policy at present under way should be very cautious, as we are operating in a fast-changing landscape.

## 5. CONCLUSION

Data-sharing policies are coming progressively into effect, and are intended to reform the way clinical research is performed by moving toward a global research community in which sharing deidentified data becomes the norm (18). Our analysis of one of the earliest experiences in the Annals suggests that things are not that simple. While it is hard to extrapolate these findings directly to the new ICMJE policy, our results highlight the need to assess whether this new policy will achieve its intended effects. In our opinion, the ICMJE policy should be assessed in terms of impact with a dedicated evaluation component, which is currently lacking in the formulation of the policy.

**Acknowledgment:** The authors would like to thank Bruno Falissard and Etienne Dantan.

**Financial Support:** This work was supported by the Fondation pour la Recherche Médicale [grant number 6616]; the Région Bretagne [Boost ERC grant 18HC432-01N]; and the French National Research Agency – ANR [Reproducibility in Therapeutic Research / ReITheR: ANR-17-CE36-0010-01].

**Declarations of interest:** None.

**Reproducible Research Statement:** Study protocol, statistical code, data set and web-appendixes are directly available on the OSF platform (<https://osf.io/8pj5e/>).

author statement

**Claude Pellen:** Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Funding acquisition. **Laura Caquelin:** Investigation, Writing - review & editing. **Alexia Jouvance-Le Bail:** Investigation, Writing - review & editing. **Jeanne Gaba:** Investigation, Writing - review & editing. **Mathilde Vérin:** Investigation, Writing - review & editing. **David Moher:** Conceptualization, Methodology, Writing - review & editing. **John P. A. Ioannidis:** Conceptualization, Methodology, Writing - review & editing. **Florian Naudet:** Conceptualization, Methodology, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

## 6. REFERENCES

1. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. Vaux DL, editor. *PLOS Biology*. 2016 Jan 4;14(1):e1002333.
2. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials. *BMJ*. 2017 Jun 5;j2372.
3. 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 Feb 13;k400.
4. Laine C, Goodman SN, Griswold ME, Sox HC. Reproducible Research: Moving toward Research the Public Can Really Trust. *Ann Intern Med*. 2007 Mar 20;146(6):450.
5. Griswold M, Laine C, Mulrow C, Schaeffer MB, Stack C. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. *Peer Review Congress*; 2013.
6. Laine C, Berkwits M, Mulrow C, Schaeffer MB, Griswold M, Goodman S. Reproducible Research: Biomedical Researchers' Willingness to Share Information to Enable Others to Reproduce Their Results. *Peer Review Congress*; 2009.
7. Goldacre B, Drysdale H, Dale A, Milosevic I, Slade E, Hartley P, et al. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. *Trials* [Internet]. 2019 Dec [cited 2019 Sep 6];20(1). Available from: <https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3173-2>
8. The Editors. Discrepancies Between Prespecified and Reported Outcomes. *Annals of Internal Medicine*. 2016 Mar 1;164(5):374.
9. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. *Science*. 2015 Jun 26;348(6242):1422–5.
10. Coady SA, Mensah GA, Wagner EL, Goldfarb ME, Hitchcock DM, Giffen CA. Use of the National Heart, Lung, and Blood Institute Data Repository. *New England Journal of Medicine*. 2017 May 11;376(19):1849–58.
11. Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ, et al. Reanalyses of Randomized Clinical Trial Data. *JAMA*. 2014 Sep 10;312(10):1024.
12. Vaduganathan M, Nagarur A, Qamar A, Patel RB, Navar AM, Peterson ED, et al. Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation*. 2018 Feb 27;137(9):938–47.
13. Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. *Clinical Trials*. 2018 Aug;15(4):413–23.
14. Tannenbaum S, Ross JS, Krumholz HM, Desai NR, Ritchie JD, Lehman R, et al. Early Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators. *Annals of Internal Medicine*. 2018 Oct 16;169(8):586.



15. Ross JS, Ritchie JD, Finn E, Desai NR, Lehman RL, Krumholz HM, et al. Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ Open*. 2016 Sep;6(9):e012769.
16. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. *Nature Human Behaviour* [Internet]. 2017 Jan [cited 2019 Sep 6];1(1). Available from: <http://www.nature.com/articles/s41562-016-0021>
17. 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 Apr 5;357:j1390.
18. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data Sharing Statements for Clinical Trials — A Requirement of the International Committee of Medical Journal Editors. *N Engl J Med*. 2017 Jun 8;376(23):2277–9.

Journal Pre-proof

	Overall	No intent to share	Intent to share	p
<b>Number of articles</b>	185	79	106	
<b>Year of publication</b> median [IQR]	2012 [2010, 2015]	2012 [2010, 2014.5]	2012 [2009, 2015]	0.797*
<b>Geographical zones</b> n (%)				0.318†
Asia	13 (7.0)	8 (10.1)	5 (4.7)	
Europe	54 (29.2)	20 (25.3)	34 (32.1)	
Other	21 (11.4)	7 (8.9)	14 (13.2)	
USA	97 (52.4)	44 (55.7)	53 (50.0)	
<b>Intervention type</b> n (%)				0.620†
Complex intervention	89 (48.1)	35 (44.3)	54 (50.9)	
Device	8 (4.3)	3 (3.8)	5 (4.7)	
Drug	88 (47.6)	41 (51.9)	47 (44.3)	
<b>Control group</b> n (%) Ref: inactive	140 (75.7)	61 (77.2)	79 (74.5)	0.804‡
<b>Medical specialty</b> n (%) Ref: surgery	4 (2.2)	3 (3.8)	1 (0.9)	0.314†
<b>Primary outcome</b> n (%) Ref: positive	123 (66.5)	58 (73.4)	65 (61.3)	0.117‡
<b>Funding source</b> n (%)				0.200†
Academic	79 (42.7)	31 (39.2)	48 (45.3)	
Charity	10 (5.4)	3 (3.8)	7 (6.6)	
Industry	28 (15.1)	17 (21.5)	11 (10.4)	
Mixed	68 (36.8)	28 (35.4)	40 (37.7)	
<b>Sample size</b> median [IQR]	299 [158, 719]	267 [151, 663]	310 [170.25, 722.75]	0.442*
<b>Protocol shared</b> n (%) Ref: yes	151 (81.6)	51 (64.6)	100 (94.3)	<0.001‡
<b>Code shared</b> n (%) Ref: yes	134 (72.4)	43 (54.4)	91 (85.8)	<0.001‡
<b>Number of citations</b> median [IQR]	59 [29, 114]	71 [36, 116.5]	51.5 [25, 112.75]	0.082*

Table 1. Characteristics of the randomized controlled trials included

\*: Kruskal-Wallis test

†: Fisher exact test

‡: Pearson's Chi-squared test

## Figures

Figure 1

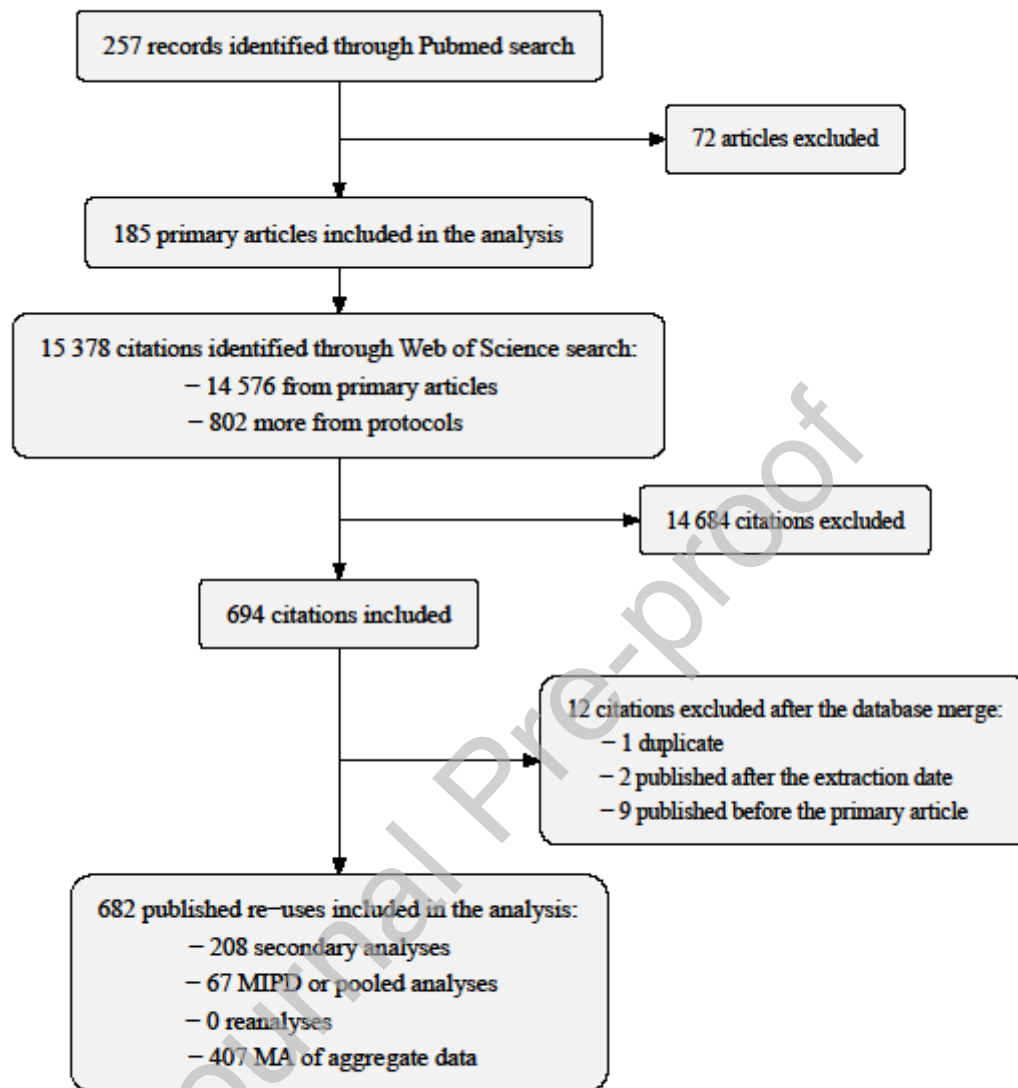


Figure 1. Study flow diagram

Figure 2

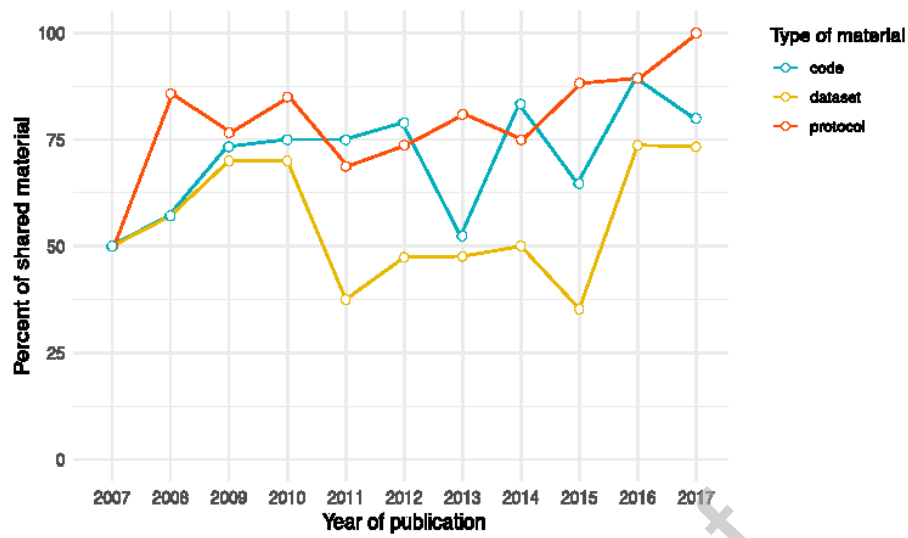
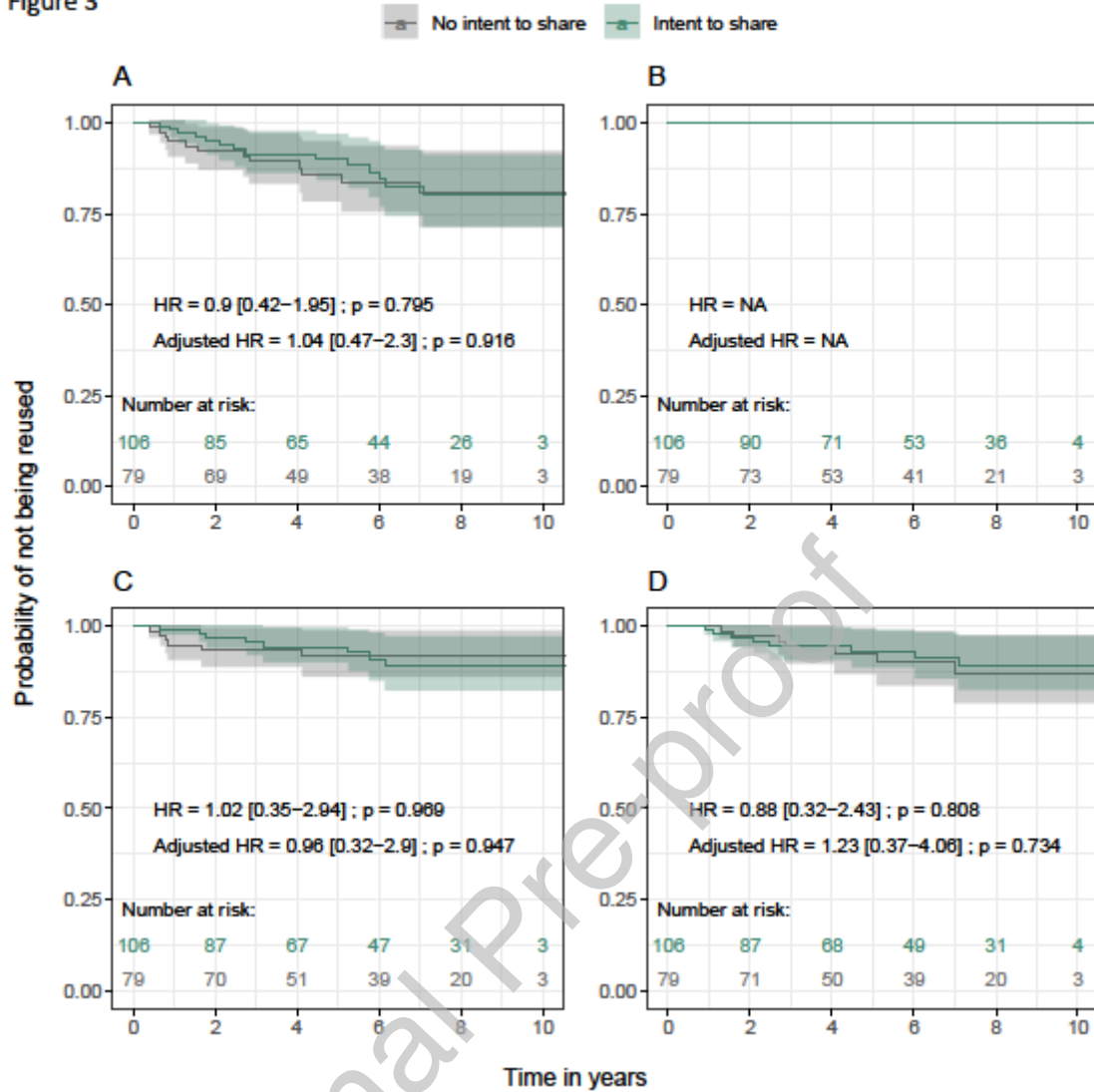


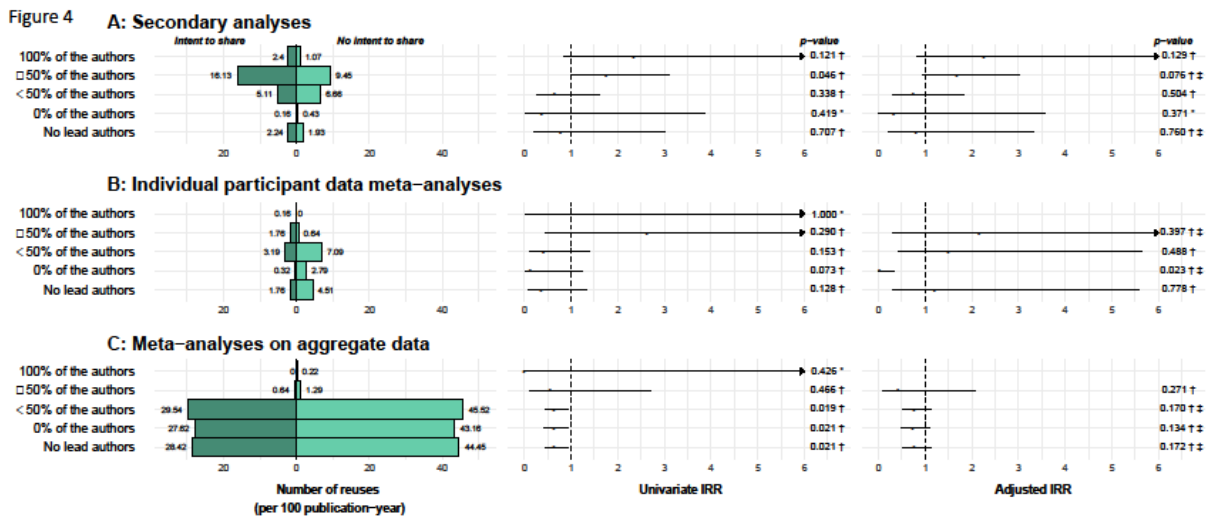
Figure 2. Data-sharing rates for protocol, statistical code and dataset of RCTs published in the Annals of Internal Medicine over time

Figure 3



**Figure 3. Kaplan-Meier curves and hazard ratios (HR) of published re-uses where the first, last and corresponding authors are not among the authors of the primary article**

*A: any type of published re-uses. B: published re-analyses. C: published secondary analyses. D: published MIPD.*



**Figure 4. Number of re-uses by intent to share**

\*: *Poisson model*

†: *Binomial negative model*

‡: *Mixed model*

### Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.