Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy

Florian Naudet, **Charlotte Sakarovitch, Perrine Janiaud, Ioana Cristea, Daniele Fanelli, David Moher, John P.A. Ioannidis**

PLOS BIOLOGY

There is a growing movement to encourage reproducibility and transparency practices in the scientific community, including public access to raw data and protocols, the conduct of replication studies, systematic integration of evidence in systematic reviews, and the documentation of funding and potential conflicts of interest. In this survey, we assessed the current status of reproducibility and transparency addressing these indicators in a random sample of 441 biomedical journal articles published in 2000–2014. Only one study provided a full protocol and none made all raw data directly available. Replication studies were rare ($n = 4$), and only 16 studies had their data included in a subsequent systematic review or meta-analysis. The majority of studies did not mention anything about funding or conflicts of interest. The percentage of articles with no statement of conflict decreased substantially between 2000 and 2014 (94.4% in 2000 to 34.6% in 2014); the percentage of articles reporting statements of conflicts (0% in 2000, 15.4% in 2014) or no conflicts (5.6% in 2000, 50.0% in 2014) increased. Articles published in journals in the clinical medicine category versus other fields were almost twice as likely to not include any information on funding and to have private funding. This study provides baseline data to compare future progress in improving these indicators in the scientific literature.

META-RESEARCH ARTICLE

Reproducible Research Practices and Transparency across the Biomedical Literature

Shareen A. Iqbal[1☯], Joshua D. Wallach[2,3☯], Muin J. Khoury[4,5], Sheri D. Schully[4], John P. A. Ioannidis[2,3,6,7*]

Medical journals can be a leverage.

Annals of Internal Medicine

EDITORIAL

Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors

Darren B. Taichman, MD, PhD, Secretary, ICMJE, Executive Deputy Editor, *Annals of Internal Medicine*
Joyce Backus, MSLS, Representative and Associate Director for Library Operations, National Library of Medicine
Christopher Baethge, MD, Chief Scientific Editor, *Deutsches Ärzteblatt (German Medical Journal)*
Howard Bauchner, MD, Editor-in-Chief, *JAMA (Journal of the American Medical Association)* and the JAMA Network
Peter W. de Leeuw, MD, Editor-in-Chief, *Nederlands Tijdschrift voor Geneeskunde (The Dutch Medical Journal)*
Jeffrey M. Drazen, MD, Editor-in-Chief, *New England Journal of Medicine*
John Fletcher, MB, BChir, MPH, Editor-in-Chief, *Canadian Medical Association Journal*
Frank A. Frizelle, MBChB, FRACS, Editor-in-Chief, *New Zealand Medical Journal*
Trish Groves, MBBS, MRCPsych, Head of Research, *British Medical Journal*
Abraham Haileamlak, MD, Editor-in-Chief, *Ethiopian Journal of Health Sciences*
Astrid James, MBBS, Deputy Editor, *The Lancet*
Christine Laine, MD, MPH, Editor-in-Chief, *Annals of Internal Medicine*
Larry Peiperl, MD, Chief Editor, *PLOS Medicine*
Anja Pinborg, MD, Scientific Editor-in-Chief, *Ugeskrift for Laeger (Danish Medical Journal)*
Peush Sahni, MBBS, MS, PhD, Representative and Past President, World Association of Medical Editors
Sinan Wu, MD, Representative, *Chinese Medical Journal*

Feedback may be posted at www.icmje.org by 18 April 2016.

Annals of Internal Medicine
ESTABLISHED IN 1927 BY THE AMERICAN COLLEGE OF PHYSICIANS

Annals of Internal Medicine

EDITORIAL

Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors

The International Committee of Medical Journal Editors (ICMJE) believes that there is **an ethical obligation** to responsibly share data generated by interventional clinical trials because **participants have put themselves at risk**.

In a growing consensus, many funders around the world—foundations, government agencies, and industry—now mandate data sharing. Here we outline ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors

Darren B. Taichman, MD, PhD, Secretary, ICMJE, Executive Deputy Editor, *Annals of Internal Medicine*
Joyce Backus, MSLS, Representative and Associate Director for Library Operations, National Library of Medicine
Christopher Baethge, MD, Chief Scientific Editor, *Deutsches Ärzteblatt (German Medical Journal)*
Howard Bauchner, MD, Editor-in-Chief, *JAMA (Journal of the American Medical Association)* and the JAMA Network
Peter W. de Leeuw, MD, Editor-in-Chief, *Nederlands Tijdschrift voor Geneeskunde (The Dutch Medical Journal)*
Jeffrey M. Drazen, MD, Editor-in-Chief, *New England Journal of Medicine*
John Fletcher, MB, BChir, MPH, Editor-in-Chief, *Canadian Medical Association Journal*
Frank A. Frizelle, MBChB, FRACS, Editor-in-Chief, *New Zealand Medical Journal*
Trish Groves, MBBS, MRCPsych, Head of Research, *British Medical Journal*
Abraham Haileamlak, MD, Editor-in-Chief, *Ethiopian Journal of Health Sciences*
Astrid James, MBBS, Deputy Editor, *The Lancet*
Christine Laine, MD, MPH, Editor-in-Chief, *Annals of Internal Medicine*
Larry Peiperl, MD, Chief Editor, *PLOS Medicine*
Anja Pinborg, MD, Scientific Editor-in-Chief, *Ugeskrift for Laeger (Danish Medical Journal)*
Peush Sahni, MBBS, MS, PhD, Representative and Past President, World Association of Medical Editors
Sinan Wu, MD, Representative, *Chinese Medical Journal*

Feedback may be posted at www.icmje.org by 18 April 2016.

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

Table. Examples of Data Sharing Statements That Fulfill These ICMJE Requirements*

	Example 1	Example 2	Example 3	Example 4
Will individual participant data be available (including data dictionaries)?	Yes	Yes	Yes	No
What data in particular will be shared?	All of the individual participant data collected during the trial, after deidentification.	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).	Not available
What other documents will be available?	Study Protocol, Statistical Analysis Plan, Informed Consent Form, Clinical Study Report, Analytic Code	Study Protocol, Statistical Analysis Plan, Analytic Code	Study Protocol	Not available
When will data be available (start and end dates)?	Immediately following publication. No end date.	Beginning 3 months and ending 5 years following article publication.	Beginning 9 months and ending 36 months following article publication.	Not applicable
With whom?	Anyone who wishes to access the data.	Researchers who provide a methodologically sound proposal.	Investigators whose proposed use of the data has been approved by an independent review committee ("learned intermediary") identified for this purpose.	Not applicable
For what types of analyses?	Any purpose.	To achieve aims in the approved proposal.	For individual participant data meta-analysis.	Not applicable
By what mechanism will data be made available?	Data are available indefinitely at (*Link to be included).*	Proposals should be directed to xxx@yyy. To gain access, data requestors will need to sign a data access agreement. Data are available for 5 years at a third party website (*Link to be included).*	Proposals may be submitted up to 36 months following article publication. After 36 months the data will be available in our University's data warehouse but without investigator support other than deposited metadata. Information regarding submitting proposals and accessing data may be found at (*Link to be provided*).	Not applicable

* These examples are meant to illustrate a range of, but not all, data sharing options.

The NEW ENGLAND JOURNAL of MEDICINE

THE LANCET

JAMA®
The Journal of the American Medical Association

JAMA Internal Medicine

PLOS | MEDICINE

thebmj

Annals of Internal Medicine
www.annals.org Established in 1927 by the American College of Physicians

BMC Medicine

No policy, never :

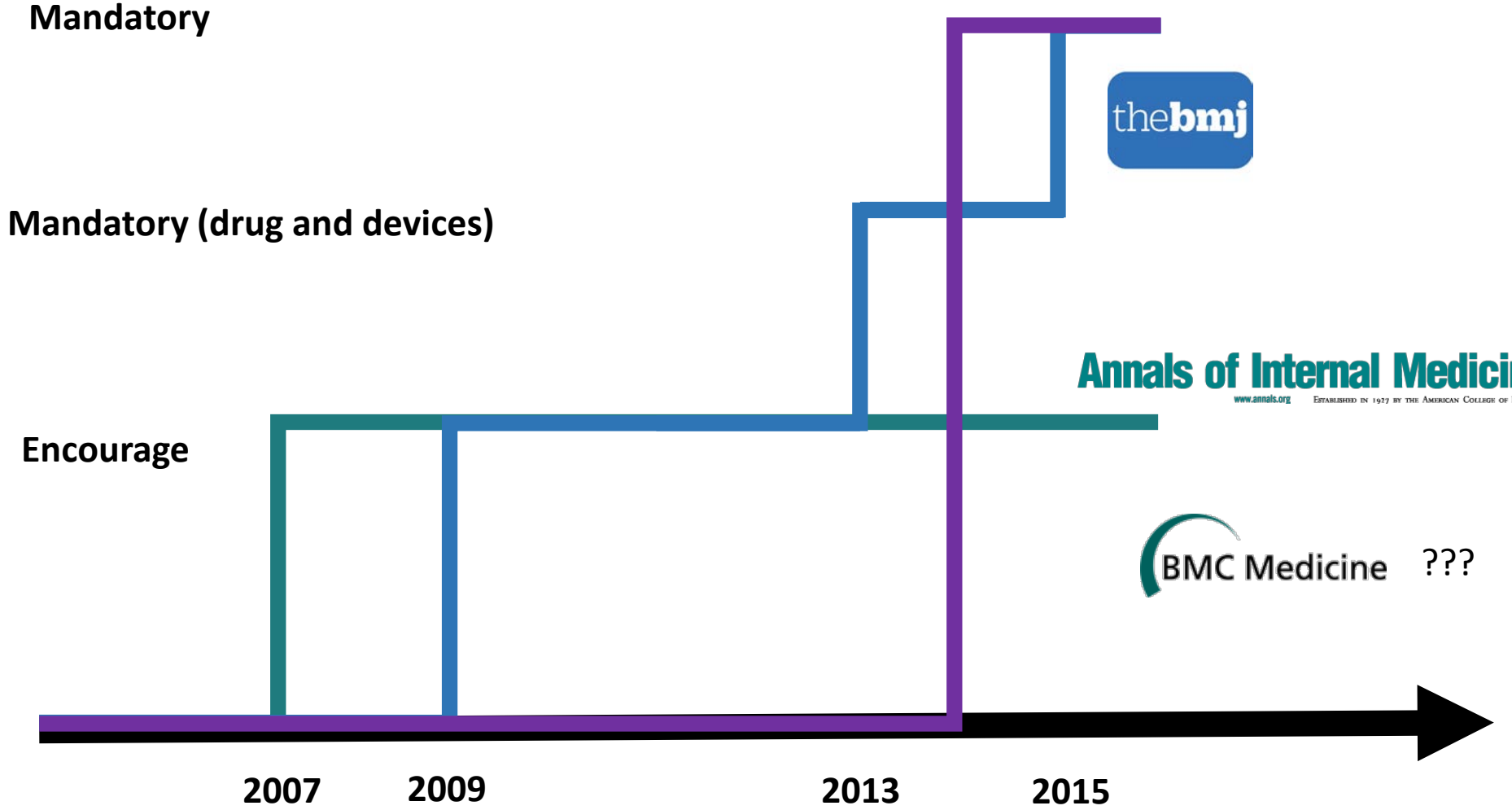
The NEW ENGLAND JOURNAL of MEDICINE

THE LANCET

JAMA®
The Journal of the American Medical Association

JAMA Internal Medicine

PLOS | MEDICINE



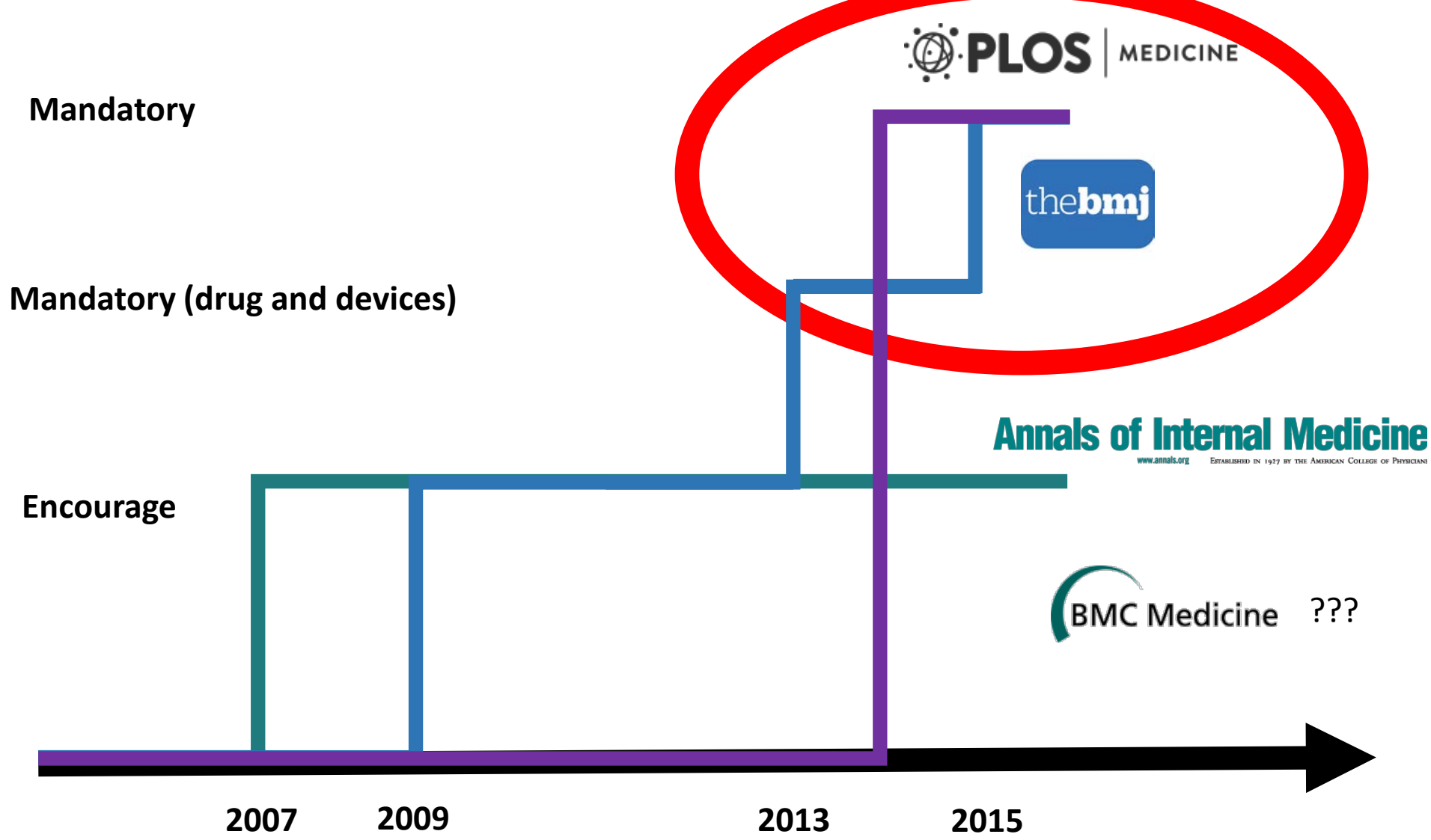
No policy, never :

The NEW ENGLAND JOURNAL of MEDICINE

THE LANCET

JAMA
The Journal of the American Medical Association

JAMA Internal Medicine



No policy, never :

The NEW ENGLAND JOURNAL of MEDICINE

THE LANCET

JAMA
The Journal of the American Medical Association

JAMA Internal Medicine

We proposed to survey all RCTs published in these 2 journals and to explore data availability and to perform re-analyses of the primary outcomes.

osf.io/u6hcv

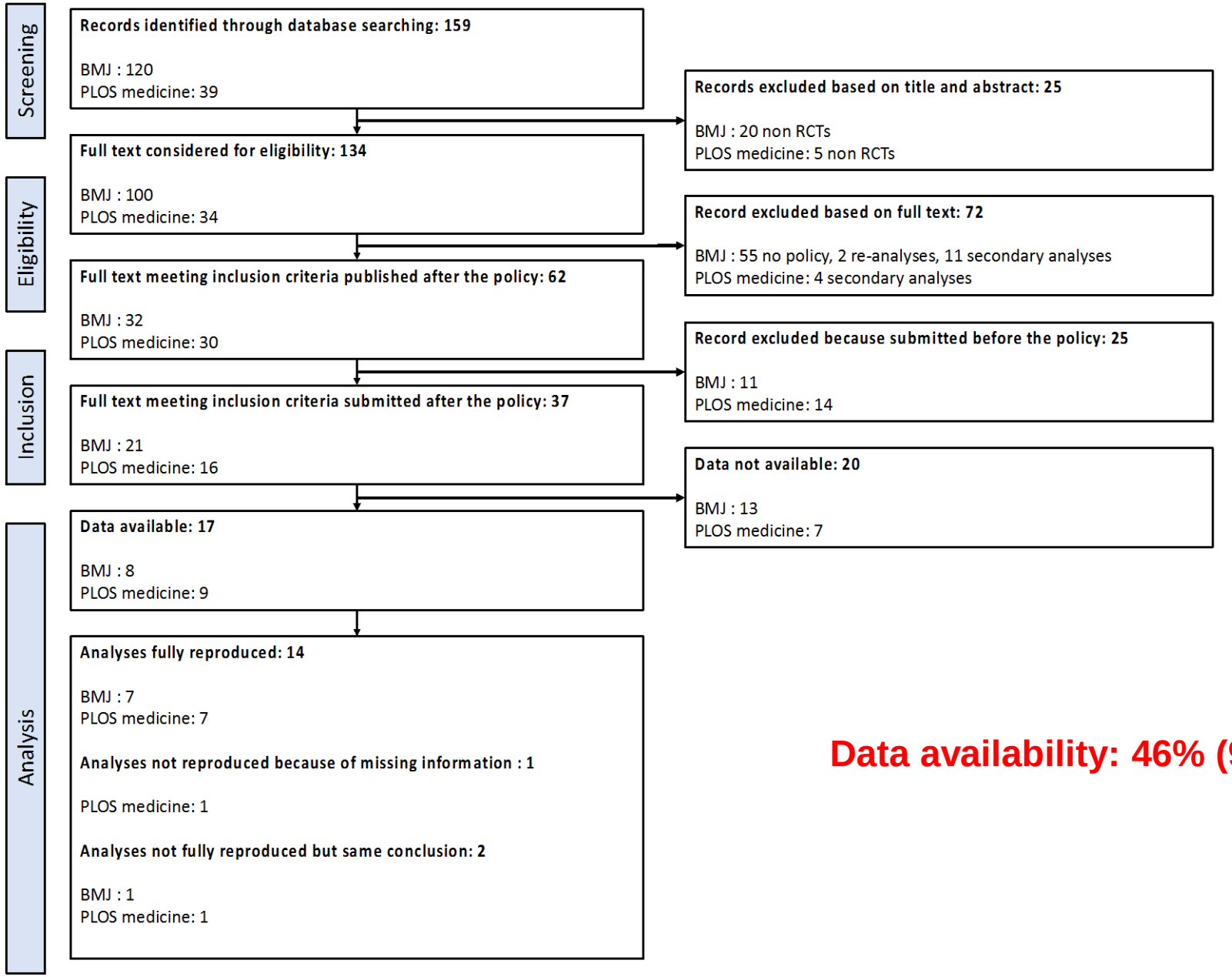
Type of outcome	Brief description
Data availability (Primary outcome)	Successful data sharing defined as the availability of data and information to allowing for a reanalysis of the primary outcomes
Type of data sharing	Accessibility of data: - Upon request by e-mail (see appendix) - Upon request on a specific website - Upon request on a specific register - Available on a public register - Other (specify)
Delay for collecting the data	- Delay in days
Reason for non availability in case data were not shared	- Privacy concerns - Technical - Non willingness to engage in sharing data - Other (specify)
De-identification of data	- Name (YES NO) - Birthdate (YES NO) - Address (YES NO)
Type of data shared [2]	- Analysable - Edited/cleaned - Computerised - Coded - Abstracted - Uncoded
Sharing of analysis code	- Yes - Yes, after a specific request - No

Type of outcome	Brief description
Data availability (Primary outcome)	Successful data sharing defined as the availability of data and information to allowing for a reanalysis of the primary outcomes
Type of data sharing	Accessibility of data: - Upon request by e-mail (see appendix) - Upon request on a specific website - Upon request on a specific register - Available on a public register - Other (specify)
Delay for collecting the data	- Delay in days
Reason for non availability in case data were not shared	- Privacy concerns - Technical - Non willingness to engage in sharing data - Other (specify)
De-identification of data	- Name (YES NO) - Birthdate (YES NO) - Address (YES NO)
Type of data shared [2]	- Analysable - Edited/cleaned - Computerised - Coded - Abstracted - Uncoded
Sharing of analysis code	- Yes - Yes, after a specific request - No

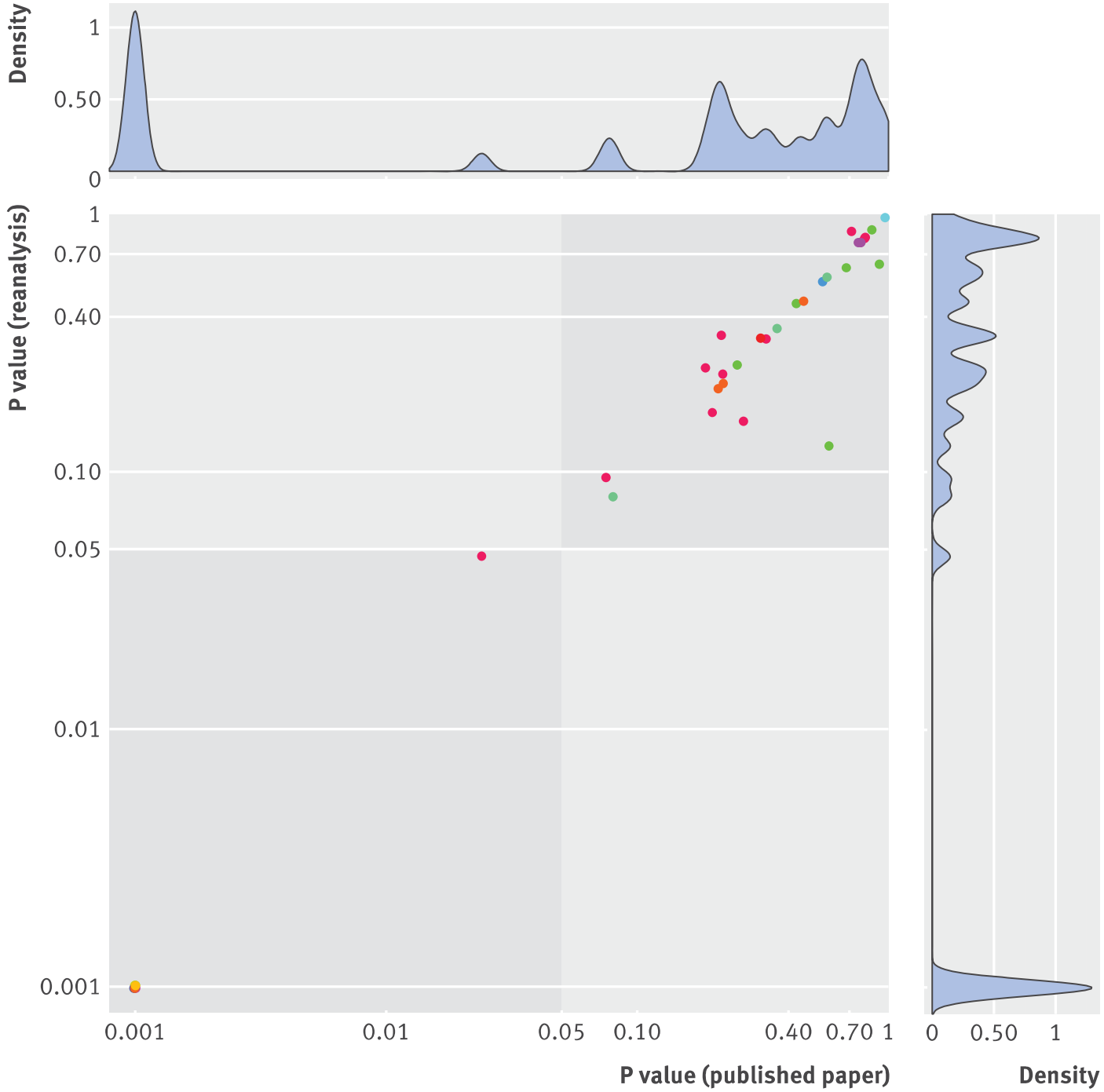
Type of outcome	Brief description
Result of a re-analysis	Result of the study reanalysis: - Not possible (with reason) - Finding not reproduced but same conclusion - Finding not reproduced but with a different conclusion - Finding reproduced

We shall also note whether the sharing of data and/or code required clarifications for which additional queries had to be made to the authors in order to obtain the relevant data and/or code, clarify their labels and/or use, and reproduce the original analysis of the primary outcomes.

A catalogue of these required clarifications will be created and similar clarifications will be grouped for descriptive purposes. The project will thus allow to generate a list of what are some common challenges and may help address these challenges pre-emptively in future published trials.



Data availability: 46% (95% CI [30% to 62%])



Analyses fully reproduced:

82%, 95% CI [59% to 94%])

Of the remaining RCTs, errors were identified in two but reached similar conclusions.

One paper **did not provide enough information in the Methods section** to reproduce the analyses

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Geographical area of the lead country			
Europe	25 (67 %)	17 (80 %)	8 (50 %)
Australia and New Zealand	4 (11 %)	1 (5 %)	3 (19 %)
Northern America	3 (8 %)	1 (5 %)	2 (12.5 %)
Africa	3 (8 %)	1 (5 %)	2 (12.5 %)
East Asia	1 (3 %)	0 (0 %)	1 (6 %)
Middle East	1 (3 %)	1 (5 %)	0 (0 %)
Type of intervention			
Drug	20 (54 %)	13 (62 %)	7 (44 %)
Device	8 (22 %)	8 (38 %)	0 (0 %)
Complex intervention	9 (24 %)	0 (0 %)	9 (56 %)
Medical specialty			
Infectious disease	12 (33 %)	4 (19 %)	8 (50 %)
Rheumatology	5 (14 %)	5 (24 %)	0 (0 %)
Endocrinology/nutrition	4 (11 %)	1 (5 %)	3 (19 %)
Paediatrics	3 (8 %)	2 (9 %)	1 (6 %)
Mental health / addiction	2 (5 %)	1 (5 %)	1 (6 %)
Obstetrics	2 (5 %)	1 (5 %)	1 (6 %)
Emergency medicine	2 (5 %)	2 (9 %)	0 (0 %)
Geriatrics	2 (5 %)	0 (0 %)	2 (13 %)
Other	5 (14 %)	5 (24 %)	0 (0 %)

Table 1: Characteristics of the included studies.

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Geographical area of the lead country			
Europe	25 (67 %)	17 (80 %)	8 (50 %)
Australia and New Zealand	4 (11 %)	1 (5 %)	3 (19 %)
Northern America	3 (8 %)	1 (5 %)	2 (12.5 %)
Africa	3 (8 %)	1 (5 %)	2 (12.5 %)
East Asia	1 (3 %)	0 (0 %)	1 (6 %)
Middle East	1 (3 %)	1 (5 %)	0 (0 %)
Type of intervention			
Drug	20 (54 %)	13 (62 %)	7 (44 %)
Device	8 (22 %)	8 (38 %)	0 (0 %)
Complex intervention	9 (24 %)	0 (0 %)	9 (56 %)
Medical specialty			
Infectious disease	12 (33 %)	4 (19 %)	8 (50 %)
Rheumatology	5 (14 %)	5 (24 %)	0 (0 %)
Endocrinology/nutrition	4 (11 %)	1 (5 %)	3 (19 %)
Paediatrics	3 (8 %)	2 (9 %)	1 (6 %)
Mental health / addiction	2 (5 %)	1 (5 %)	1 (6 %)
Obstetrics	2 (5 %)	1 (5 %)	1 (6 %)
Emergency medicine	2 (5 %)	2 (9 %)	0 (0 %)
Geriatrics	2 (5 %)	0 (0 %)	2 (13 %)
Other	5 (14 %)	5 (24 %)	0 (0 %)

Table 1: Characteristics of the included studies.

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Designs			
Superiority (Head to head)	18 (49 %)	15 (71 %)	3 (19 %)
Superiority (Factorial)	1 (3 %)	1 (5 %)	0 (0 %)
Superiority (Clusters)	8 (21 %)	1 (5 %)	7 (43 %)
Non-inferiority + Superiority (Head to head)	4 (11 %)	1 (5 %)	3 (19 %)
Non-inferiority (Head to head)	6 (16 %)	3 (14 %)	3 (19 %)
Sample size	432 (213 – 1070)†	221 (159 – 494)	1047 (433 – 2248)†
Private sponsorship			
No	26 (70 %)	15 (71 %)	11 (69 %)
Provided the device	1 (3 %)	1 (5 %)	0 (0 %)
Provided the intervention	1 (3 %)	0 (0 %)	1 (6 %)
Provided the drug	5 (13 %)	1 (5 %)	4 (25 %)
Provided the drug and some financial support	2 (5 %)	2 (9 %)	0 (0 %)
Provided partial financial support	1 (3 %)	1 (5 %)	0 (0 %)
Provided total financial support	1 (3 %)	1 (5 %)	0 (0 %)
Statement of availability			
Ask to contact by e-mail	23 (62 %)	17 (81 %)	6 (38 %)
Explain how to retrieve the data (e.g. platform)	9 (24 %)	0 (0 %)	9 (56 %)
State ‘no additional data available‘	2 (5 %)	2 (9 %)	0 (0 %)
Ask to contact by mail	1 (3 %)	0 (0 %)	1 (6 %)
Embargo	1 (3 %)	1 (5 %)	0 (0 %)
No statement	1 (3 %)	1 (5 %)	0 (0 %)

Table 1: Characteristics of the included studies.

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Designs			
Superiority (Head to head)	18 (49 %)	15 (71 %)	3 (19 %)
Superiority (Factorial)	1 (3 %)	1 (5 %)	0 (0 %)
Superiority (Clusters)	8 (21 %)	1 (5 %)	7 (43 %)
Non-inferiority + Superiority (Head to head)	4 (11 %)	1 (5 %)	3 (19 %)
Non-inferiority (Head to head)	6 (16 %)	3 (14 %)	3 (19 %)
Sample size	432 (213 – 1070)†	221 (159 – 494)	1047 (433 – 2248)†
Private sponsorship			
No	26 (70 %)	15 (71 %)	11 (69 %)
Provided the device	1 (3 %)	1 (5 %)	0 (0 %)
Provided the intervention	1 (3 %)	0 (0 %)	1 (6 %)
Provided the drug	5 (13 %)	1 (5 %)	4 (25 %)
Provided the drug and some financial support	2 (5 %)	2 (9 %)	0 (0 %)
Provided partial financial support	1 (3 %)	1 (5 %)	0 (0 %)
Provided total financial support	1 (3 %)	1 (5 %)	0 (0 %)
Statement of availability			
Ask to contact by e-mail	23 (62 %)	17 (81 %)	6 (38 %)
Explain how to retrieve the data (e.g. platform)	9 (24 %)	0 (0 %)	9 (56 %)
State ‘no additional data available‘	2 (5 %)	2 (9 %)	0 (0 %)
Ask to contact by mail	1 (3 %)	0 (0 %)	1 (6 %)
Embargo	1 (3 %)	1 (5 %)	0 (0 %)
No statement	1 (3 %)	1 (5 %)	0 (0 %)

Table 1: Characteristics of the included studies.

Perceived costs/benefit ratio of re-analyses

Costs involved in the data-sharing process

“As the study was launched we did not plan the cost for this preparation”

“*Took some time to translate it […] the original one was done in Hebrew*”

“[We] decided to do the work for free although this is some extra work. For future projects it will be important to consider these costs either on your [side] or in the grant application for the trials”

Perceived benefits of sharing data for the purpose of this study

“*We could create such a dataset […] but it would require substantial effort and we cannot do it simply to demonstrate that it is possible.*”

“We are especially keen that our data are used for IPD meta-analyses and have shared this with […] we see that as an exemplar of meaningful data-sharing. Yours is a most unusual request”

“A slight concern about ‘naming and shaming’ individual studies/investigators”

Novelty and heterogeneity in data-sharing practices

Some authors who were unsure how to proceed

"[…] However, I am just wanting to confirm School policy and our ethical obligations regarding the sharing of data before we proceed"

"*Please can you let me know how you have been receiving data from other centers securely?*"

Heterogeneity between different procedures to share data

Open repository (n=5)

Downloadable on a secured website (n=1) after registration

Included as appendix of the published paper (n=3)

Sent by e-mail (n=10).

In 3 occasions, we signed a data-sharing request/agreement. In addition, typically there was no standard in type of data-shared. In one case, authors mentioned explicitly that they followed standardized guidelines[15] to prepare the dataset.

Incomplete or ambiguous labels and reporting

Complexity of some analyses

Obtaining more information about the analytic method by contacting authors was sometimes (6 studies) necessary

Incomplete information

Three databases did not provide sufficient information to reproduce the analyses:

- Variables used for adjustment
- Definition of the analysis population
- Randomization groups

Communication with authors was therefore necessary and was fruitful in one these 3 cases.

Limitations

… 2 very selected journals.

Reproducibility of their analysis VS an analysis with the best standards

Availability of data is a surrogate of data sharing policies

Reproducibility in therapeutic research

ReiTheR

Reproductibilité de la recherche thérapeutique

Task	Conduct	Deliverable
1	2018	- A detailed description of all platforms enabling data sharing, including their content; - A detailed description of current editorial policies; - A detailed description of various funder policies; - A specific description of these practices in France;
2	2018-2019	- An estimation of reproducibility of statistical analyses of RCTs in therapeutic research;
3	2018-2019	- An estimation of reproducibility of statistical analyses of RCTs in EMA approvals;
4	2018-2020	- An estimation of reproducibility of safety analyses of RCTs in therapeutic research;
5	2020	- Some preliminary data on effects of funding (a form of contextual sensitivity);
6	2018-2020	- A tool for scoring 1/ good practice in data sharing and 2/ reproducibility.

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

We need evidence about the interest of data sharing policies.

High quality policies must be evidence-based.

It is therefore necessary to assess whether they have the intended effects. All these initiatives must have an evaluation component which is often missing.

This is both a scientific and an ethical imperative.

We need evidence about the interest of data sharing policies.

High quality policies must be evidence-based.

It is therefore necessary to assess whether they have the intended effects. All these initiatives must have an evaluation component which is often missing.

This is both a scientific and an ethical imperative.

… ELSE DATA-SCIENTISTS WILL STILL BE CALLED PARASITES.

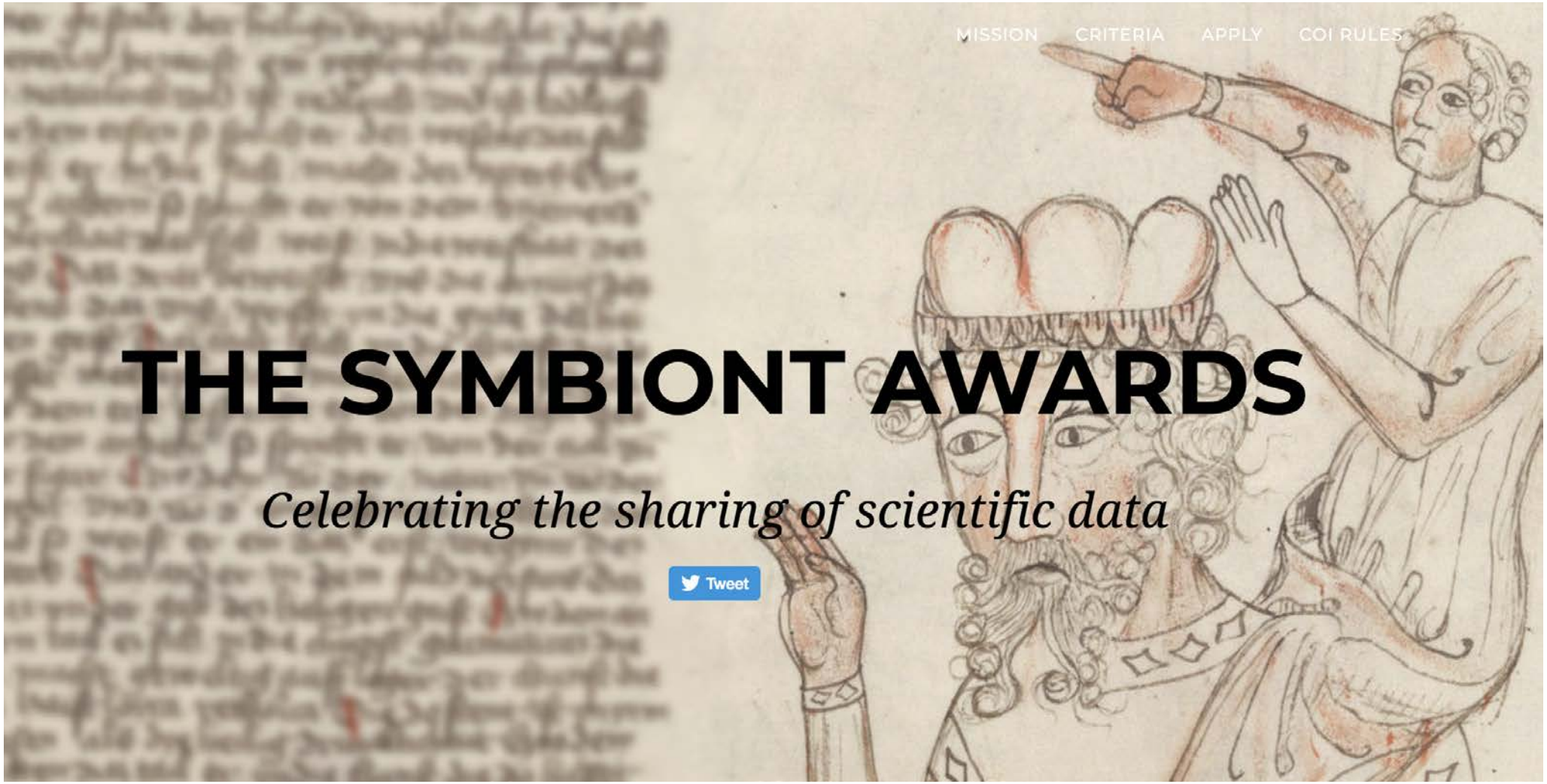
The NEW ENGLAND JOURNAL of MEDICINE

A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as "research parasites."

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.


THE PARASITE AWARDS
Celebrating rigorous secondary data analysis



Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

We need incentives.

This study was made possible through sharing of anonymized individual participant data from the authors of all studies. We thank the authors who were contacted for this study: C Bullen and the National Institute for Health Innovation, S Gilbody, C Hewitt, L Littlewood, C van der Meulen, H van der Aa, S Cohen, M Bicket, T Harris, the STOP GAP study investigators including Kim Thomas, Alan Montgomery, and Nicola Greenlaw, Nottingham University Hospitals NHS Trust, NIHR programme grants for applied research, the Nottingham Clinical Trials Unit, C Polyak, K Yuhas, C Adrion, U Mansmann, G Greisen, S Hyttel-Sørensen A Barker, R Morello, K Luedtke, M Paul, D Yahav, L Chesterton, the Arthritis Research UK Primary Care Centre, and C Hanson.

Thank you.