

## **Supplementary Material**

Supplement to the manuscript entitled “Evaluation of Clinical Trial Data Sharing Policy in Leading Medical Journals” by Valentin Danchev, Yan Min, John Borghi, Mike Baiocchi, and John P.A. Ioannidis.

## **Appendix 1: Inclusion criteria, search strategy, and data collection and analysis**

### **Inclusion and exclusion criteria**

In order to be eligible to be selected in the study, a published paper must meet the following inclusion criteria:

1. Publication reports clinical trial results
2. Published in JAMA, NEJM, Lancet
3. Published since July 1, 2018
4. Type of publication is Article
5. Contain a Data Sharing Statement (DSS)\*

\*Articles published in 2020 that meet criteria 1 to 4 are eligible even if no DSS is present as these are likely submitted after July 1, 2018, and therefore fall under the requirements of the ICMJE data sharing policy.

Excluded from the study were publications that contain no Data Sharing Statement due to:

1. Submission prior to July 1, 2018
2. Study not a clinical trial (e.g., observational)
3. Type of publication is letter/correspondence

### **Search strategy**

A MEDLINE/PubMed search was performed on April 4, 2020 using the following search strategy:

```
((("The New England journal of medicine"[Journal]) OR "JAMA"[Journal]) OR "Lancet (London, England)"[Journal]) AND clinical trial[Publication Type]) AND ("2018/07/01"[Date - Publication] : "2020/04/04"[Date - Publication])
```

As of April 04 2020, the search yielded 629 results. Out of those, 486 publications were clinical trials and contained a DSS. One 2020 publication met 1 to 4 inclusion criteria but contained no DSS, and was included in the study sample as not sharing data because articles published in 2020 were likely submitted after policy's effective date, July 1, 2018. We conducted a cross-sectional observational study of the resulting sample of 487 articles.

### **Independent review of articles**

For each of the 487 articles, two reviewers independently evaluated the Data Sharing Statement and funding statement, using procedures described in the Codebook. Discrepancies were resolved unanimously or by a third reviewer.

### **Data analysis**

We performed data analysis and generated summary statistics and data visualizations using Python (Python Software Foundation. Python Language Reference, version 3.6.8. Available at <http://www.python.org>), Jupyter Notebook,<sup>1</sup> and the following libraries: SciPy,<sup>2</sup> Pandas,<sup>3</sup> NumPy,<sup>4</sup> Matplotlib,<sup>5</sup> Scikits-Bootstrap,<sup>6</sup> Seaborn.<sup>7</sup> All computer code used in this analysis is available at <https://osf.io/s5vbg/>.

## **Appendix 2: Codebook**

### **A. Declared data sharing**

1 = YES → Code **B, C, D,** and **F**

0 = NO → Code **E** and **F**

### **B. Type of declared available data [1 if applicable, blank otherwise]**

Deidentified individual participant data

Aggregate data only

Unspecified/ partial data

### **C. Access to data [1 if applicable, blank otherwise]**

Request to authors

Request to committee/ group/ unit

Request to company

Request to repository/ archive

    Name of repository [Name as noted on DSS]

    Repository contains information about the data/study [1 = Yes, 0 = No]

    Data is available on repository [1 = Yes, 0 = No]

Access unspecified

Data is available to others

### **D. Conditions for data sharing [1 if applicable, blank otherwise]**

Embargo

    Embargo period [in Months]

    If embargo period unspecified, copy wording from the DSS

Collaboration

Product approval

### **E. Reasons for why data not available [1 if applicable, blank otherwise]**

No reason given

Data privacy

Time and cost

Ongoing trial/ research

Regulatory approval

Proprietary data

Shared among co-investigators only

Data may be available for collaboration

Data may be available upon request

### **F. Funder type\***

1 = Industry

2 = Non-industry NIH

3 = Non-industry non-NIH

4 = Mixed (any combination of industry, non-industry NIH, non-industry non-NIH)

\* Only funding is eligible (in-kind support or supplies provided at no charge are not eligible).

## Appendix 3: Detailed description of clinical trial data repositories

Table S1. Description of Clinical Trial IPD Repositories <sup>a</sup>			
Acronym / short name	Full name	URL	Host organization
CSDR	Clinical Study Data Request	<a href="https://www.clinicalstudydatarequest.com">https://www.clinicalstudydatarequest.com</a>	Consortium of clinical study sponsors/funders
NCTN/NCORP Data Archive	National Clinical Trials Network (NCTN) and NCI Community Oncology Research Program (NCORP) Data Archive	<a href="https://nctn-data-archive.nci.nih.gov">https://nctn-data-archive.nci.nih.gov</a>	National Cancer Institute (NCI)
BioLINCC	Biologic Specimen and Data Repository Information Coordinating Center	<a href="https://biolincc.nhlbi.nih.gov">https://biolincc.nhlbi.nih.gov</a>	National Heart, Lung, and Blood Institute (NHLBI)
YODA	The Yale University Open Data Access (YODA) Project	<a href="https://yoda.yale.edu/">https://yoda.yale.edu/</a>	Yale University
Vivli	Center for Global Clinical Research Data	<a href="https://vivli.org">https://vivli.org</a>	The Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard
DASH	NICHD Data and Specimen Hub	<a href="https://dash.nichd.nih.gov/">https://dash.nichd.nih.gov/</a>	The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)
NINDS Archived Clinical Research Datasets	NINDS Archived Clinical Research Datasets	<a href="https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets">https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets</a>	National Institute of Neurological Disorders and Stroke (NINDS)
BORIS	Bern Open Repository and Information System	<a href="https://boris.unibe.ch/">https://boris.unibe.ch/</a>	University of Bern and the Bern University Hospital
NIDDK Central Repository	NIDDK Central Repository	<a href="https://repository.niddk.nih.gov">https://repository.niddk.nih.gov</a>	National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
JCHR	JAEB Center for Health Research	<a href="https://www.jaeb.org/">https://www.jaeb.org/</a>	JAEB Center for Health Research

<sup>a</sup> Included are only repositories that were referenced in two or more data sharing statements.

## References

1. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows. 2016; 2016. p. 87–90.
2. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 2020; **17**(3): 261–72.
3. McKinney W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference; 2010: Austin, TX; 2010. p. 51–6.
4. Walt Svd, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng* 2011; **13**(2): 22–30.
5. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* 2007; **9**(3): 90–5.
6. Evans CG. Scikits-Bootstrap (Version v1.0.1). Zenodo. <http://doi.org/10.5281/zenodo.3548989>; 2019 November 20.
7. Waskom M, Botvinnik O, Ostblom J, et al. mwaskom/seaborn: v0.10.1 (April 2020) (Version v0.10.1). Zenodo. <http://doi.org/10.5281/zenodo.3767070>; 2020, April 26.