

Implementing a data sharing culture

C. Barbui*, O. Gureje, B. Puschner, S. Patten and G. Thornicroft, *EPS Editors*

First published online 6 May 2016

Data sharing is the process of sharing with other researchers the deidentified individual patient data underlying the results presented in scientific articles. Recently, the US Institute of Medicine urged biomedical journals, as evaluators and publishers of research results, and implementers of academic standards, to enforce policies that require sharing of clinical trial data (Institute of Medicine, 2015). The International Committee of Medical Journal Editors (ICMJE) pointed out that there is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk (Longo & Drazen, 2016; Taichman *et al.* 2016). In addition to clinical trial data, it has been argued that results from observational studies should similarly be shared for the same reasons (Barbui, 2016).

Accountability is a first reason for sharing data. By accessing the raw information underlying the results presented in an article, other researchers may re-run the same analyses that are presented by the study authors, or may plan different analyses to answer the same research question, thus confirming (replicating) the main findings or raising concerns about their robustness and validity under different analytical or statistical assumptions. A second reason for sharing the data is that other researchers may use the shared dataset to answer a different research question. A third reason is that shared datasets of similarly collected data may be used within systematic reviews, to run meta-analyses and individual-patient data meta-analyses. Some also argue that shared datasets may be effectively used for educational purposes (Feldman *et al.* 2012).

There are also challenges in sharing data (Gewin, 2016), such as preserving patient privacy and confidentiality (Sarpawari *et al.* 2014), giving credit to those who conducted the study and collected the data (Longo & Drazen, 2016), the cost for developing reliable repositories of data, and the additional work and cost for

researchers, who may be asked to develop datasets suitable for use by others and perhaps to pay for hosting the data in a repository. Running new analyses on shared datasets may also pose scientific issues, as these new analyses, by definition, cannot be pre-planned and may therefore suffer from being guided by the data.

However, solutions to these challenges are coming. Researchers should consider the issue of data sharing from the very beginning of their research projects, by including for example a data sharing plan in the study protocol. It may report a procedure for including provision for data sharing when gaining informed consent, preserving patient confidentiality and privacy when data are shared (Sarpawari *et al.* 2014; El Emam *et al.* 2015), and details of what is planned to be shared and how, with a timeline. This is relevant as developing a high-quality database in a way suitable for secondary uses may require considerable work, and how the raw data are shared may depend on the type of data that are collected (sometimes it is possible to present the raw data in the main manuscript or in additional supporting files, sometimes a web-based repository is needed). Researchers should also draft a detailed publication plan, with a timeline, in order to give those who gathered the data a chance to make the best use of the database (Drazen, 2014).

Current policies of biomedical journals are very heterogeneous (Barbui, 2016). Some journals do not mention data sharing, and do not require any statement to be published along with the study report on the possibility to access the raw data. Other journals encourage data sharing, and require a formal statement describing under which conditions raw data are accessible. A third policy is implemented by PLOS journals, which require full availability of all data underlying the findings described in published study reports.

Epidemiology and Psychiatric Sciences announces the implementation of the following requirement on data sharing:

“For all research articles (randomised controlled trials, observational studies, systematic reviews and meta-analyses) authors are encouraged to link their articles to the raw data from their studies. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available

* Address for correspondence: Professor C. Barbui, Department of Public Health and Community Medicine, Section of Psychiatry, Ospedale Policlinico ‘G.B. Rossi’, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Italy
(Email: corrado.barbui@univr.it)

and appropriate) or presented in the main manuscript or additional supporting files whenever possible. All authors must include an 'Availability of Data and Materials' section in their manuscript detailing where the data supporting their findings can be found. Authors who do not wish to share their data must state that data will not be shared, and give the reason."

We hope to contribute to the implementation of a new data sharing culture. Currently, a scientific output only corresponds to a study report published in a medical journal, while in the near future it might consist of all materials described in the manuscript, including all relevant raw data. We need to change how we think about data (Drazen, 2015). Data sharing and open science are the future of science.

Acknowledgement

None.

Financial Support

None.

Conflict of Interest

No financial conflicts of interest to declare. CB, as one of the Editors of the Cochrane Common Mental Disorders group, has a strong interest in accessing

raw data from clinical studies. As a researcher, CB has recently been involved in the publication of an observational study that required full data sharing.

References

- Barbui C** (2016). Sharing all types of clinical data and harmonizing journal standards. *BMC Med* **14**, 63.
- Drazen JM** (2014). Open data. *New England Journal of Medicine* **370**, 662.
- Drazen JM** (2015). Sharing individual patient data from clinical trials. *New England Journal of Medicine* **372**, 201–202.
- El Emam K, Rodgers S, Malin B** (2015). Anonymising and sharing individual patient data. *BMJ* **350**, h1139.
- Feldman L, Patel D, Ortmann L, Robinson K, Popovic T** (2012). Educating for the future: another important benefit of data sharing. *Lancet* **379**, 1877–1878.
- Gewin V** (2016). Data sharing: an open mind on open data. *Nature* **529**, 117–119.
- Institute of Medicine** (2015). Sharing clinical trial data: maximizing benefits, minimizing risk. <http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk>.
- Longo DL, Drazen JM** (2016). Data sharing. *New England Journal of Medicine* **374**, 276–277.
- Sarpawari A, Kesselheim AS, Malin BA, Gagne JJ, Schneeweiss S** (2014). Ensuring patient privacy in data sharing for postapproval research. *New England Journal of Medicine* **371**, 1644–1649.
- Taichman DB, Backus J, Baethge C, Bauchner H, Leeuw PW, Drazen JM, Fletcher J, Frizelle FA** (2016). Sharing clinical trial data. *BMJ* **532**, i255.