

JHP POLICY ON DATA SHARING

David F Marks

Editor, JHP

editorjhp@gmail.com

For a trial period of one year, from 1 July 2018, JHP will be requiring the following as conditions of consideration for publication of all empirical reports of quantitative studies in the journal. The new rules apply to studies using quantitative but not qualitative data and in mixed-methods studies will apply only to the quantitative parts of the study report.

Please note these new data sharing principles will be applied to *all* empirical studies, not only clinical trial report data.

- 1) As of 1 July 2018, manuscripts concerning clinical trials and other empirical studies that are submitted to JHP must contain a Data Sharing Statement as described below.
- 2) Any clinical trial that begins enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial's registration.

The author will be required to share only those data described in the publication. To ensure access, the data should be uploaded to a suitable data-sharing repository such as one listed in the Registry of Research Data Repositories:

<http://www.re3data.org/>

A link to the repository should be included at the end of the paper. For registered clinical trials, if the data sharing plan changes after registration, this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.

Data Sharing Statements must indicate the following: whether individual de-identified participant data (including data dictionaries) will be shared; what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.); when the data will become available and for how long; by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism).

The data to be shared should be useable and interpretable:

1. If the data are in the form of a **statistical dataset**, variables must be labelled clearly, and variables that are stored as labelled numeric values must have associated value labels. The version of the software used to create the dataset must be stipulated (to clarify potential back-compatibility issues).
2. For data stored as a **spreadsheet, or delimited text**, an associated text file containing variable labels and, where appropriate, value labels for labelled numeric data.
3. **Missing data codes** should be documented, together with numbers of missing values for each variable. Ideally, missing data should be left blank, not assigned a pseudo-numeric code.
4. **Measurement units** and **measurement times** (where appropriate).
5. The dataset should be accompanied by a **codebook** giving means of continuous variables and frequencies of categorical variables, together with numbers of valid cases. This allows the use to check that they have read the data correctly into whatever software they are using.

An example of a Data Sharing Statement that would meet these requirements is as follows:

Will de-identified individual participant data be available (including data dictionaries)? Yes

What data in particular will be shared? All of the individual participant data collected during the study, after de-identification.

What other documents will be available? Study Protocol, Statistical Analysis Plan, Informed Consent Form, Clinical Study Report, Analytic Code.

When will data be available (start and end dates)? Immediately following publication. No end date.

With whom? Anyone who wishes to access the data.

For what types of analyses? Any purpose.

By what mechanism will data be made available? Data are available indefinitely at (Link to an appropriate be included).

The Data Sharing Statement should appear at the end of the Method section of the manuscript.

The new rules will be reviewed after one year to determine their feasibility over the longer term.