Data sharing: issues and available guidance

EQUATOR Network Seminar, 3rd October 2011

Iain Hrynaszkiewicz
Journal Publisher, BioMed Central
iain.hrynaszkiewicz@biomedcentral.com
Low transparency

- Trial registration
- Protocol publication
- Protocol amendments
- Methodology articles
- First report of trial findings

Potential for bias

- Secondary analysis
- Longer-term follow-up
- Lessons learned
- Systematic review and meta-analysis

High transparency

- Data publication, deposition and sharing
Data sharing – the six W’s

- Why
- Where
- What
- How
- When
- Who
Why share data

- Testing of additional hypotheses
- Teaching
- Validation of previous findings
- Integration with other data sets
- Simplification and enhancement of future systematic reviews and meta-analyses

Why share data (cont)

- Reduction of error and fraud
- Increased academic credit (citations)
- Economic benefits
- Reduce waste/duplication of effort
- Drives new knowledge discovery
- Ultimately, better patient care

Avoiding reporting bias

October 2010: Reboxetine is “overall an ineffective and potentially harmful antidepressant”
Reporting bias in medical research - a narrative review

Natalie McGauran*, Beate Wieseler, Julia Kreis, Yvonne-Beatrice Schüler, Heike Kölsch and Thomas Kaiser

Abstract

Reporting bias represents a major problem in the assessment of health care interventions. Several prominent cases have been described in the literature, for example, in the reporting of trials of antidepressants, Class I anti-arhythmic drugs, and selective COX-2 inhibitors. The aim of this narrative review is to gain an overview of reporting bias in the medical literature, focusing on publication bias and selective outcome reporting. We explore whether these types of bias have been shown in areas beyond the well-known cases noted above, in order to gain an impression of how widespread the problem is. For this purpose, we screened relevant articles on reporting bias that had previously been obtained by the German Institute for Quality and Efficiency in Health Care in the context of its health technology assessment reports and other research work, together with the reference lists of these articles.

We identified reporting bias in 40 indications comprising around 50 different pharmacological, surgical (e.g. vacuum assisted closure therapy), diagnostic (e.g. ultrasound), and preventive (e.g. cancer vaccines) interventions. Regarding pharmacological interventions, cases of reporting bias were, for example, identified in the treatment of the following conditions: depression, bipolar disorder, schizophrenia, anxiety disorder, attention-deficit hyperactivity disorder, Alzheimer’s disease, pain, migraine, cardiovascular disease, gastric ulcers, irritable bowel syndrome, urinary incontinence, atopic dermatitis, diabetes mellitus type 2, hypercholesterolemia, thyroid disorders, menopausal symptoms, various types of cancer (e.g. ovarian cancer and melanoma), various types of infections (e.g. HIV, influenza and Hepatitis B), and acute trauma. Many cases involved the withholding of study data by manufacturers and regulatory agencies or the active attempt by manufacturers to suppress publication. The ascertained effects of reporting bias included the overestimation of efficacy and the underestimation of safety risks of interventions.

In conclusion, reporting bias is a widespread phenomenon in the medical literature. Mandatory prospective registration of trials and public access to study data via results databases need to be introduced on a worldwide scale. This will allow for an independent review of research data, help fulfill ethical obligations towards patients, and ensure a basis for fully-informed decision making in the health care system.
Journal policies

BioMed Central/Public Library of Science
Submission/publication implies willingness to share data/readily reproducible materials with other scientists on request.

Annals Internal Med/BMJ
Statement about availability of materials for reproducible research/data sharing required in published manuscript.

Nature
As a condition of publication supporting data must be made available to editors and peer-reviewers at the time of submission.
Walport M, Brest P: **Sharing research data to improve public health.** *The Lancet.* DOI:10.1016/S0140-6736(10)62234-9

More funder data sharing policies at: [http://biosharing.org/policies](http://biosharing.org/policies)
"Science" special issue on data sharing:
http://www.sciencemag.org/site/special/data/

Travis K, Feb 2011: Sharing Data in Biomedical and Clinical Research
http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2011_02_11/careedit.a1100014

http://bit.ly/oL8mg8

Commons Select Committee

MPs call for research data to be fully disclosed and made publicly available

28 July 2011

Report indicates that the oversight of research integrity in the UK is unsatisfactory.

The Science and Technology Committee today concludes that in order to allow others to repeat and build on experiments, researchers should aim for the gold standard of making their data fully disclosed and made publicly available.

- Report: Peer review in scientific publications
- Inquiry: Peer review in scientific publications
- Science and Technology Committee
What data to share/publish?

- Aggregated collection of patient observations used for the summary statistical findings presented in the main report of the research project
- The minimum level of detail necessary to reproduce all numbers reported

Where to publish data?

- Data included as online journal supplementary material (additional files)
- Data papers ("data notes") e.g. Trials (http://www.trialsjournal.com), BMC Research Notes
- Institution/domain-specific data repositories e.g. Edinburgh DataShare (http://datashare.is.ed.ac.uk/), Dryad (http://datadryad.org)
How to share data?
Protecting privacy

- Ethical guidelines e.g. International Committee of Medical Journal Editors, Committee on Publication Ethics
- Legal requirements e.g. HIPAA (USA), Data Protection Act (UK)
Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers

Iain Hrynaszkiewicz, Melissa L Norton, Andrew J Vickers, Douglas G Altman

**Iain Hrynaszkiewicz and colleagues** propose a minimum standard for anonymising datasets to ensure patient privacy when sharing clinical research data. Many peer-reviewed journals' instructions for authors require that authors should be prepared to share their data.

In Europe, the Data Protection Directive (Directive 95/46/EU) provides a harmonised data protection law.

**BMJ** 2010;340:c181
Co-published in: **Trials** 2010, **11**:9
**Anonymisation**

"...datasets that contain **three or more indirect identifiers**, such as age or sex, should be reviewed by an independent researcher or ethics committee”

Hrynasykiewicz *et al.*, *BMJ* 2010;340:c181

<table>
<thead>
<tr>
<th>Identifier (information sources)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td></td>
</tr>
<tr>
<td>Initials</td>
<td></td>
</tr>
<tr>
<td>Address, including full or partial postal code</td>
<td></td>
</tr>
<tr>
<td>Telephone or fax numbers or contact information</td>
<td></td>
</tr>
<tr>
<td>Electronic mail addresses</td>
<td></td>
</tr>
<tr>
<td>Unique identifying numbers</td>
<td></td>
</tr>
<tr>
<td>Vehicle identifiers</td>
<td></td>
</tr>
<tr>
<td>Medical device identifiers</td>
<td></td>
</tr>
<tr>
<td>Web or internet protocol addresses</td>
<td></td>
</tr>
<tr>
<td>Biometric data</td>
<td></td>
</tr>
<tr>
<td>Facial photograph or comparable image</td>
<td></td>
</tr>
<tr>
<td>Audiotapes</td>
<td></td>
</tr>
<tr>
<td>Names of relatives</td>
<td></td>
</tr>
<tr>
<td>Dates related to an individual (including date of birth)</td>
<td></td>
</tr>
<tr>
<td><strong>Indirect—may present risk if present in combination with others in the list</strong></td>
<td></td>
</tr>
<tr>
<td>Place of treatment or health professional responsible for care</td>
<td>Could be inferred from investigator affiliations</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Rare disease or treatment</td>
<td></td>
</tr>
<tr>
<td>Sensitive data, such as illicit drug use or “risky behaviour”</td>
<td></td>
</tr>
<tr>
<td>Place of birth</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic data, such as occupation or place of work, income, or education</td>
<td>MRC requirement is for “rare” occupations only</td>
</tr>
<tr>
<td>Household and family composition</td>
<td></td>
</tr>
<tr>
<td>Anthropometry measures</td>
<td></td>
</tr>
<tr>
<td>Multiple pregnancies</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Small denominators—population size of &lt;100</td>
<td></td>
</tr>
<tr>
<td>Very small numerators—event counts of &lt;3</td>
<td></td>
</tr>
<tr>
<td>Year of birth or age (this article)</td>
<td>Age is potentially identifying if the recruitment period is short and is fully described</td>
</tr>
<tr>
<td>Verbatim responses or transcripts</td>
<td></td>
</tr>
</tbody>
</table>
File preparation

- Data should be clean and well-annotated
- Include statistical code if possible
- Use open (e.g. XML, RAW) and common proprietary file formats (e.g. Excel, Stata) if applicable/available
- If possible, adhere to established data/metadata standards e.g. 
  http://www.cdisc.org/,
  http://www.biosharing.org/standards
Use in practice

BMJ Instructions for authors:
http://resources.bmj.com/bmj/authors/types-of-article/research

“We also strongly support the view that researchers should seek informed consent to data sharing from research participants …… Consent is particularly important because participants may be identifiable in a dataset - even an "anonymised" one that does not contain names or addresses.

The combination of three or more indirect identifiers such as age, sex, and an unusual clinical detail may be enough for at least the participant, or another interested party, to recognise themselves.”
Use in practice

The International Stroke Trial database

Peter AG Sandercock1∗, Maciej Niewada2,3, Anna Członkowska2,3 and the International Stroke Trial Collaborative Group

∗Corresponding author: Peter AG Sandercock Peter.Sandercock@ed.ac.uk

1Department of Clinical Neurosciences, University of Edinburgh, Department of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XH, UK
2Department of Clinical and Experimental Pharmacology, Warsaw Medical University, Poland, Krakowskie Przedmiescie 26/28, 00-927 Warsaw, Poland
32nd Department of Neurology, Institute of Psychiatry and Neurology, 9 Sobieskiego, 02-957 Warsaw, Poland

For all author emails, please log on.


The electronic version of this article is the complete one and can be found online at:
http://www.trialsjournal.com/content/12/1/101
“We aimed to make individual patient data from the International Stroke Trial (IST), one of the largest randomised trials ever conducted in acute stroke, available for public use, to facilitate the planning of future trials and to permit additional secondary analyses.”
When to share data?

- Let’s be pragmatic
- Data later is better than data never
- Enable retrospective data publication in recognition of its value
- Built in “temporal latencies” established in genomics community

Contreras, J L: Prepublication Data Release, Latency, and Genome Commons, Science, 2010 10.1126/science.1189253
Whose data are they anyway?

- Is it morally right for researchers or sponsors to keep patient data?
- Many opportunities from participant ownership
- Patients sharing data associated with improved health outcomes

Terry SF and Terry PF: *Power to the People: Participant Ownership of Clinical Trial Data*  
Copyright and licenses

• Publishers should not require transfer of copyright for datasets published as supplementary material
• Data/facts are *usually* not copyrightable
• Removal of intellectual property in data maximises potential for reuse, integration and new knowledge discovery
The data should be released in standardized formats without intellectual property constraints.

Conclusions

• Rather than ‘why share data?’, the question is ‘how’?
• Data sharing is a means to make research more effective
• We can better serve the health of patients with transparency
Questions?

EQUATOR Network Seminar, 3rd October 2011

Iain Hrynaszkiewicz
Journal Publisher, BioMed Central
iain.hrynaszkiewicz@biomedcentral.com