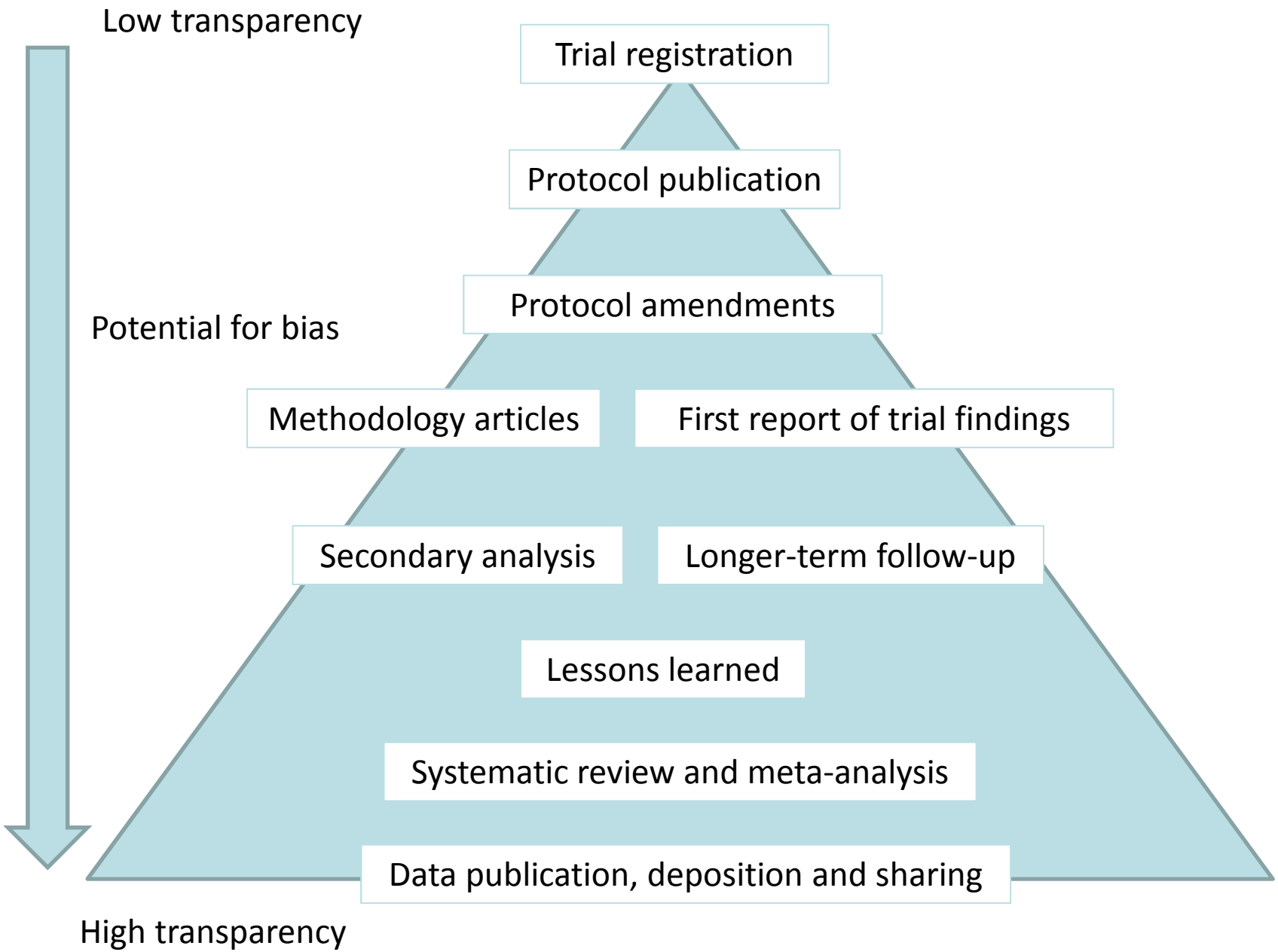
Data sharing: issues and available guidance

EQUATOR Network Seminar, 3rd October 2011

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

BioMed Central
The Open Access Publisher


Low transparency
Potential for bias
High transparency
Trial registration
Protocol publication
Protocol amendments
Methodology articles
First report of trial findings
Secondary analysis
Longer-term follow-up
Lessons learned
Systematic review and meta-analysis
Data publication, deposition and sharing

Data sharing – the six W's

- **W**hy
- **W**here
- **W**hat
- Ho**W**
- **W**hen
- **W**ho

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

Vickers, AJ: **Whose data set is it anyway? Sharing raw data from randomized trials**. *Trials* 2006, **7**:15

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

Hrynaszkiewicz I, Altman DG: **Towards agreement on best practice for publishing raw clinical trial data**. *Trials* 2009, **10**:17

Avoiding reporting bias

October 2010: Reboxetine is "overall an ineffective and potentially harmful antidepressant"

BMJ 2010;341:c4737 doi:10.1136/bmj.c4737 (Published 12 October 2010)
Cite this as: BMJ 2010;341:c4737

Research

Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials


OPEN ACCESS

Dirk Eyding, project manager[1] , **Monika Lelgemann**, senior researcher[2] , **Ulrich Grouven**, statistician[3 4] , **Martin Härter**, head of department of medical psychology[5] , **Mandy Kromp**, statistician[3] , **Thomas Kaiser**, head of department of drug assessment[3] , **Michaela F Kerekes**, data manager[3] , **Martin Gerken**, researcher[6] , **Beate Wieseler**, deputy head of department of drug assessment[3]

+ Author Affiliations

Correspondence to: B Wieseler, Institute for Quality and Efficiency in Health Care, Dillenburger Strasse 27, 51105 Cologne, Germany
beate.wieseler@iqwig.de

McGauran *et al. Trials* 2010, **11**:37
http://www.trialsjournal.com/content/11/1/37

REVIEW **Open Access**

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-arrhythmic drugs, and selective COX-2 inhibitors. The aim of this narrative review is to gain an overview of reporting bias in the medical literature, focussing 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, hypercholesterolaemia, 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 fulfil 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*

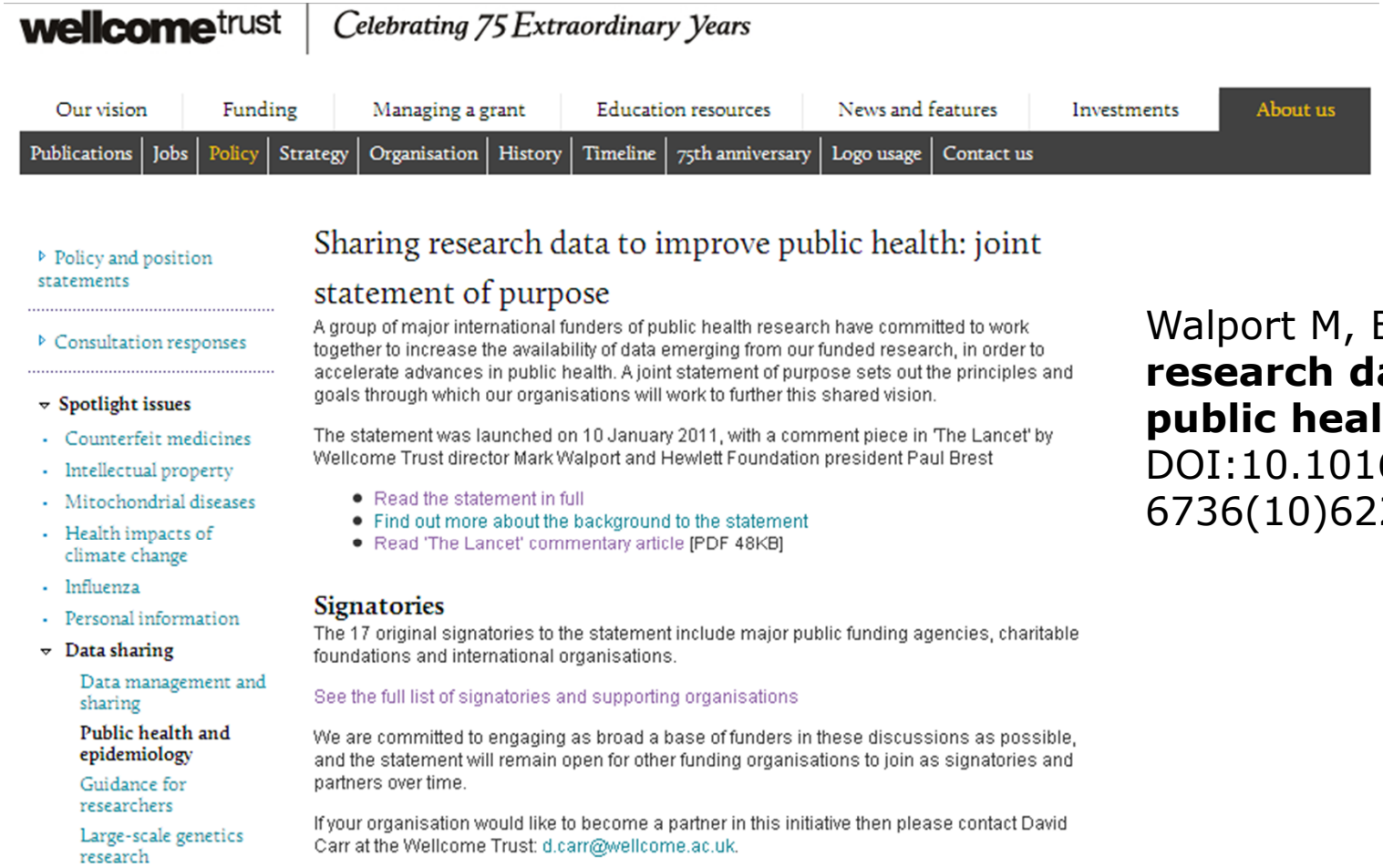
Annals Internal Med/BMJ

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*

http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030689.htm



Walport M, Brest P: **Sharing research data to improve public health**. *The Lancet*. DOI:10.1016/S0140-6736(10)62234-9

BioMed Central
The Open Access Publisher

More funder data sharing policies at:
http://biosharing.org/policies

http://bit.ly/oL8mg8

***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/caredit.a1100014

BioMed Central
The Open Access Publisher

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

Hrynaszkiewicz *et al.*: **Preparing raw clinical data for publication: guidance for authors, editors and peer reviewers**.
BMJ 2010;340:c18 // *Trials* 2010, **11**:9

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)

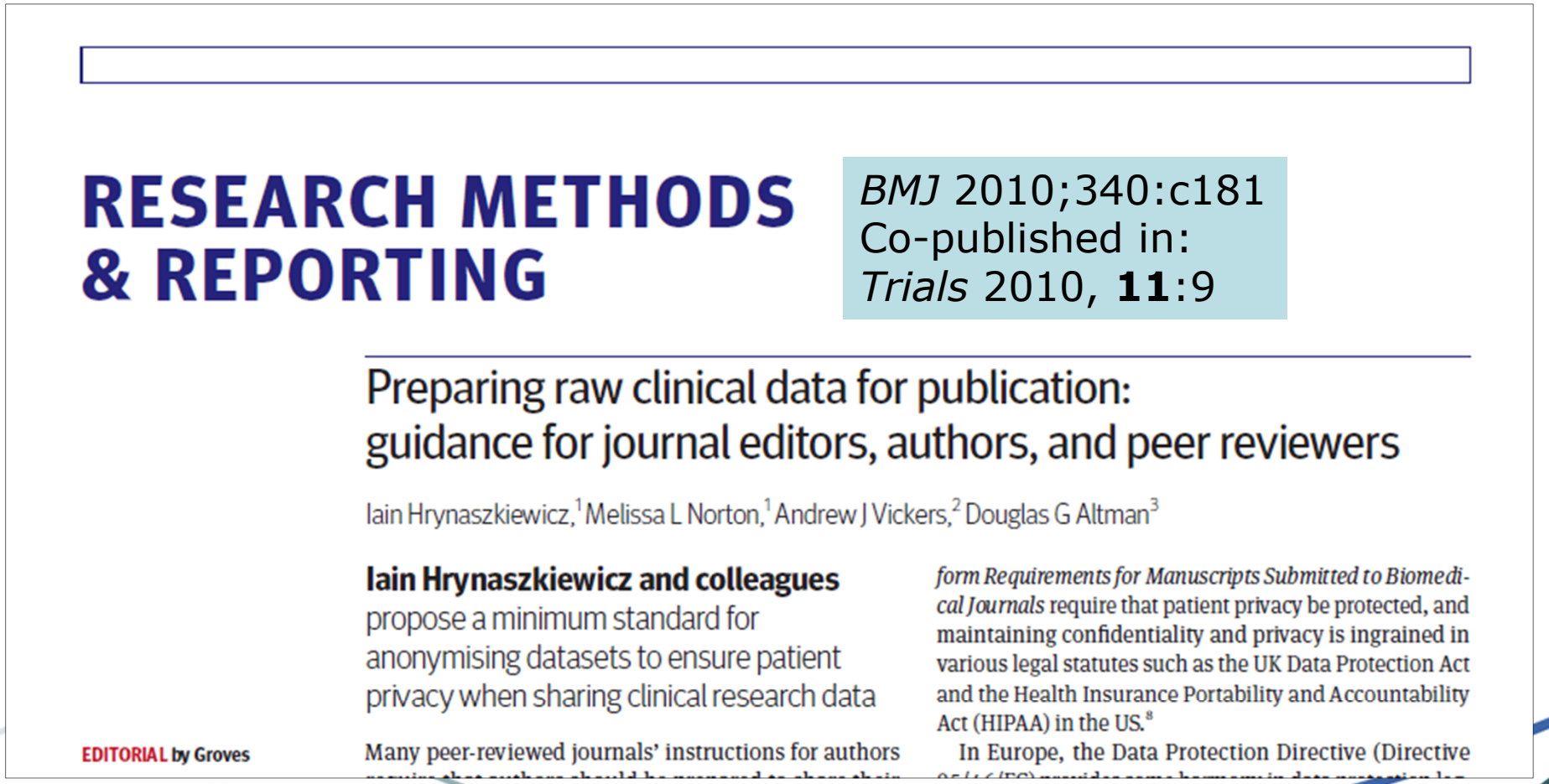
HoW to share data?

BioMed Central
The Open Access Publisher

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)

HoW to share data?



RESEARCH METHODS & REPORTING

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

Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers

Iain Hrynaszkiewicz,[1] Melissa L Norton,[1] Andrew J Vickers,[2] Douglas G Altman[3]

Iain Hrynaszkiewicz and colleagues propose a minimum standard for anonymising datasets to ensure patient privacy when sharing clinical research data

EDITORIAL by Groves

Many peer-reviewed journals' instructions for authors

form Requirements for Manuscripts Submitted to Biomedical Journals require that patient privacy be protected, and maintaining confidentiality and privacy is ingrained in various legal statutes such as the UK Data Protection Act and the Health Insurance Portability and Accountability Act (HIPAA) in the US.[8]

In Europe, the Data Protection Directive (Directive

BioMed Central
The Open Access Publisher

Anonymisation

List of potential patient identifiers in datasets

Identifier (information sources)	Comments
Direct	
Name [8 15]	
Initials [13]	
Address, including full or partial postal code [8-15]	
Telephone or fax numbers or contact information [8 10 12 15]	
Electronic mail addresses [8]	
Unique identifying numbers [8-15]	Generalised HIPAA items 7-10, 18
Vehicle identifiers [8]	
Medical device identifiers [8]	
Web or internet protocol addresses [8]	
Biometric data [8]	
Facial photograph or comparable image [8 10 11 13]	
Audiotapes [11]	
Names of relatives [10]	
Dates related to an individual (including date of birth) [8 9 11 15]	
Indirect—may present a risk if present in combination with others in the list	
Place of treatment or health professional responsible for care [10 15]	Could be inferred from investigator affiliations
Sex [9]	
Rare disease or treatment [10]	
Sensitive data, such as illicit drug use or "risky behaviour" [15]	
Place of birth [10 15]	
Socioeconomic data, such as occupation or place of work, income, or education [9 10 12 15]	MRC requirement is for "rare" occupations only
Household and family composition [15]	
Anthropometry measures [15]	
Multiple pregnancies [15]	
Ethnicity [9]	
Small denominators—population size of <100 [14]	
Very small numerators—event counts of <3 [14]	
Year of birth or age (this article)	Age is potentially identifying if the recruitment period is short and is fully described
Verbatim responses or transcripts [15]	

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

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

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

TRIALS IMPACT FACTOR 2.08

Search this journal for Go

Advanced Search

Home | Articles | Authors | Reviewers | About this journal | My Trials

Top
Abstract
Background
Methods
Results
Discussion
Competing interests
Note for user...
Authors' contributions
Sources of Fu...
Acknowledgements
References

This article is part of the series Sharing clinical research data.

Research

Highly accessed | Open access

The International Stroke Trial database

Peter AG Sandercock[1]*, **Maciej Niewada**[2,3], **Anna Członkowska**[2,3] and **the International Stroke Trial Collaborative Group**

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

Author Affiliations

1 Department of Clinical Neurosciences, University of Edinburgh, Department of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XU, UK
2 Department of Clinical and Experimental Pharmacology, Warsaw Medical University, Poland, Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland
3 2nd Department of Neurology, Institute of Psychiatry and Neurology, 9 Sobieskiego, 02-957 Warsaw, Poland

For all author emails, please log on.

Trials 2011, **12**:101 doi:10.1186/1745-6215-12-101

The electronic version of this article is the complete one and can be found online at: http://www.trialsjournal.com/content/12/1/101

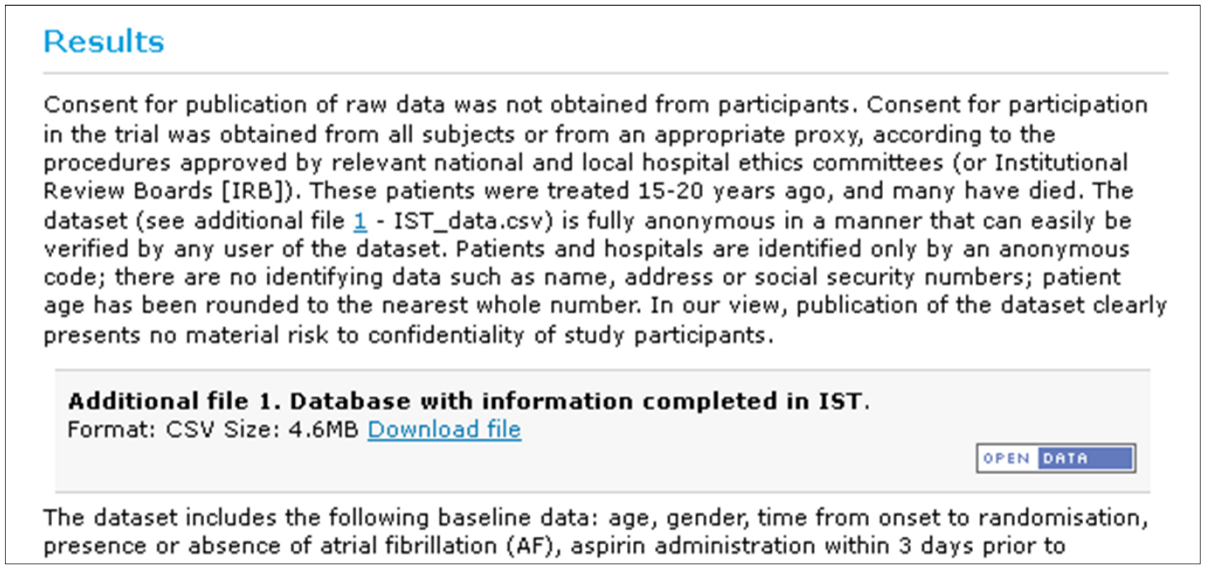
Trials
Volume 12

Viewing options
Abstract
Full text
PDF (240KB)
Additional files

Associated material
PubMed record
About this article
Readers' comments
Pre-publication history

Related literature
Articles citing this article
on Google Scholar
on ISI Web of Science
on PubMed Central
Other articles by authors
on Google Scholar
on PubMed

"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."



Results

Consent for publication of raw data was not obtained from participants. Consent for participation in the trial was obtained from all subjects or from an appropriate proxy, according to the procedures approved by relevant national and local hospital ethics committees (or Institutional Review Boards [IRB]). These patients were treated 15-20 years ago, and many have died. The dataset (see additional file 1 - IST_data.csv) is fully anonymous in a manner that can easily be verified by any user of the dataset. Patients and hospitals are identified only by an anonymous code; there are no identifying data such as name, address or social security numbers; patient age has been rounded to the nearest whole number. In our view, publication of the dataset clearly presents no material risk to confidentiality of study participants.

Additional file 1. Database with information completed in IST.
Format: CSV Size: 4.6MB Download file

OPEN DATA

The dataset includes the following baseline data: age, gender, time from onset to randomisation, presence or absence of atrial fibrillation (AF), aspirin administration within 3 days prior to

BioMed Central
The Open Access Publisher

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

BioMed Central
The Open Access Publisher

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** *Sci. Transl. Med*. 3, 69cm3 (2011).

BioMed Central
The Open Access Publisher

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

Panton Principles

Principles for Open Data in Science

Sage BIONETWORKS

Published online 20 January 2010 | Nature | doi:10.1038/news.2010.20

News

GlaxoSmithKline goes public with malaria data

Company to place structures and properties of drug leads in the public domain.

(GIGA)n SCIENCE

patientslikeme®

THAT'SMYDATA

Ball A: How to License Research Data (JISC/Digital Curation Centre), June 2011 :http://www.dcc.ac.uk/webfm_send/332

"*The data should be released in standardized formats without intellectual property constraints.*" Conway PH, VanLare JM: Improving Access to Health Care Data: The Open Government Strategy. *JAMA* 2010;**304**(9):1007-1008.

BioMed Central
The Open Access Publisher

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

BioMed Central
The Open Access Publisher