Threads, waterfalls and open access: buzzwords or bywords for modern publication planning?

December 6th 2011

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

BioMed Central
The Open Access Publisher

Open access

- There are no barriers to access (subscriptions)
- The publisher generally does not acquire any exclusive rights (e.g. Creative Commons licenses)
- Typically the publisher is paid for the service of publication
- Contrast with traditional model where research community transfers its rights to the publisher and the publisher covers costs by selling access

About BioMed Central

- Largest global publisher of peer-reviewed open access journals
- Launched first open access journals in 2000
- Now publishes over 220 open access journals
- >110,000 peer-reviewed open access articles published
- Part of Springer Science+Business Media
- Publishing benefits include visibility; speed; impact; retention of copyright; compliance with mandates

BioMed Central
The Open Access Publisher

Where is open access in 2011?

- Over 6000 open access (OA) journals in the DOAJ
- Over 1000 OA journals are indexed by Thomson Reuters (>100 from BioMed Central)
- More than 10% of the literature is published OA (2009)*
- Growth rate greater in OA publishing than non-OA
- Open access to research mandated in over 110 institutions and by nearly 50 funders
- Many publishers launching open access programs, options and “mega journals”

*Pollock D: An Open Access Primer – Market Size and Trends. Outsell inc. Vol 3, Sept 2009

BioMed Central
The Open Access Publisher

"Waterfalls"

- Cascades?
- Transfers?
- Deflections?
- Pyramids?

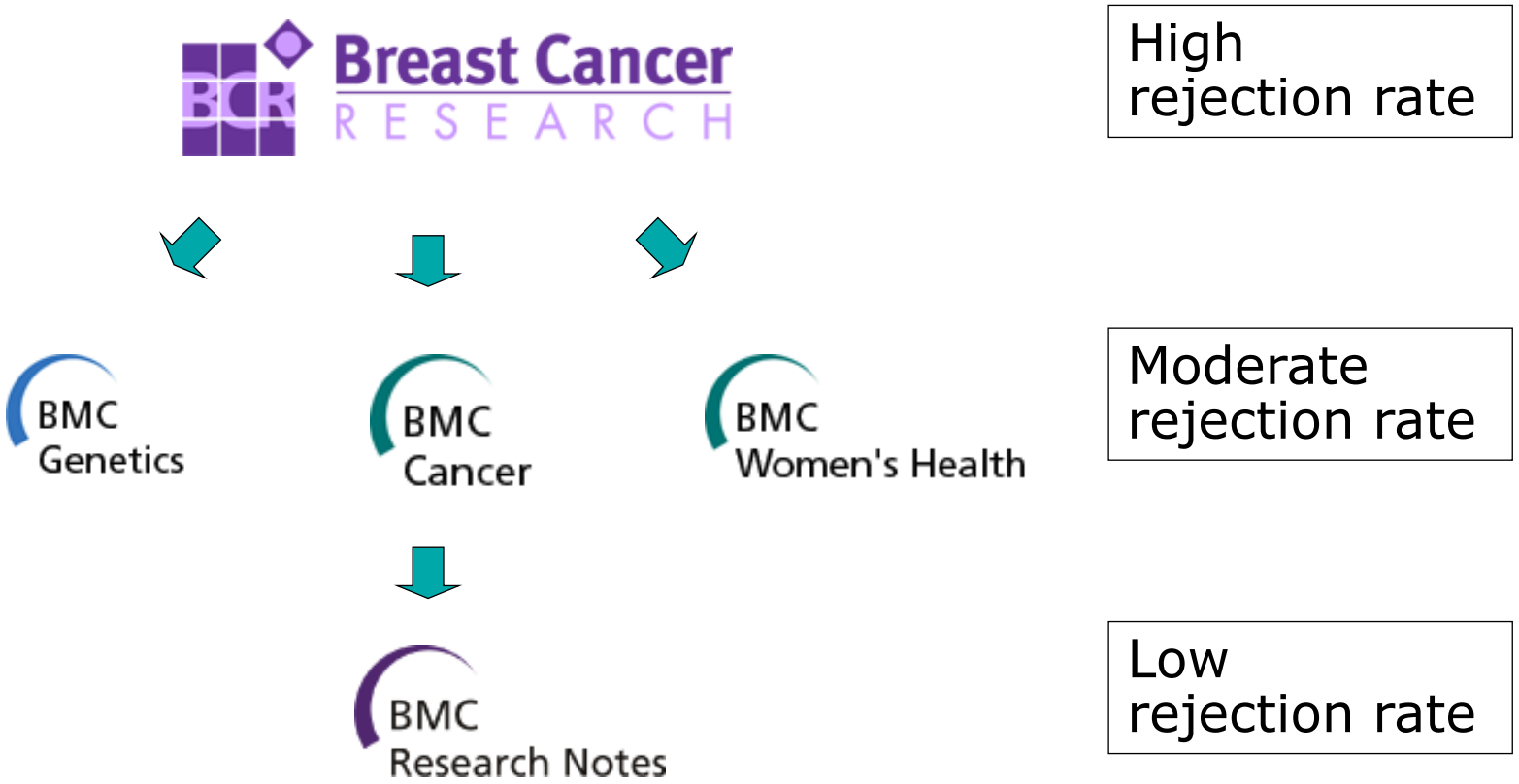
Peer review is inefficient

"*[T]he burden on researchers of reviewing papers is excessive, and we need to move away from the current system where the same paper is often reviewed multiple times by different journals.*"

- Written evidence submitted by the Wellcome Trust to Parliament
http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/856we09.htm

Unpaid peer review costs are estimated at £1.9bn globally
http://www.rin.ac.uk/system/files/attachments/Activities-costs-flows-summary.pdf

Peer review cascade



Advantages of the cascade

- Avoids delays for authors
- Avoids repeated, redundant peer review
- Separates soundness from level of interest
  - Soundness determines whether to publish
  - Interest determines where to publish
- For the publisher, high-prestige, high rejection rate titles are magnets for research articles

Peer review cascade

- Model plays an important role at BioMed Central
- Many other publishers operating similar systems e.g. PLoS (One) BMJ (Open), Nature (Scientific Reports)
- Intra-publisher and inter-publisher transfers occur e.g. Neuroscience peer review consortium
- BioMed Central developing editor tools for even more efficient transfers

http://bit.ly/oL8mg8

***Science* special issue on data sharing:**
http://www.sciencemag.org/site/special/data/

and

Data replication and reproducibility (Dec 2011):
http://www.sciencemag.org/site/special/data-rep/

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

BioMed Central and data publishing/sharing

***Trials* journal thematic series on 'Sharing clinical research data':**
http://www.trialsjournal.com/series/sharing
***BMC Research Notes* data standardization, sharing and publication:**
http://www.biomedcentral.com/bmcresnotes/series/datasharing
BMC Open Data Blog:
http://blogs.openaccesscentral.com/blogs/bmcblog/category/Open+Data

http://www.trialsjournal.com/series/sharing

Trials

BioMed Central

Editorial

Open Access

Towards agreement on best practice for publishing raw clinical trial data

Iain Hrynaszkiewicz*[1] and Douglas G Altman[2]

Address: [1]BioMed Central Ltd, 236 Gray's Inn Road, London, WC1X 8HL, UK and [2]Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford, OX2 6UD, UK

Email: Iain Hrynaszkiewicz* - iain.hrynaszkiewicz@biomedcentral.com; Douglas G Altman - doug.altman@csm.ox.ac.uk

* Corresponding author

Published: 18 March 2009

Trials 2009, **10**:17 doi:10.1186/1745-6215-10-17

Received: 3 March 2009

Accepted: 18 March 2009

This article is available from: http://www.trialsjournal.com/content/10/1/17

Abstract

Many research-funding agencies now require open access to the results of research they have funded, and some also require that researchers make available the raw data generated from that research. Similarly, the journal *Trials* aims to address inadequate reporting in randomised controlled trials, and in order to fulfil this objective, the journal is working with the scientific and publishing communities to try to establish best practice for publishing raw data from clinical trials in peer-reviewed biomedical journals. Common issues encountered when considering raw data for publication include patient privacy – unless explicit consent for publication is obtained – and ownership, but agreed-upon policies for tackling these concerns do not appear to be addressed in the guidance or mandates currently established. Potential next steps for journal editors and publishers, ethics committees, research-funding agencies, and researchers are proposed, and alternatives to journal publication, such as restricted access repositories, are outlined.

Introduction

Assessment of the reliability of published articles is seriously impeded by incomplete reporting [1]. But even if a study is impeccably reported, we usually have access only to summary information from a limited number of analyses. The availability of individual patient data, 'raw data', to the scientific community would allow many other analyses and realise a variety of benefits for science and, as a consequence, patient care. Indeed, recommendations for sharing data resulting from publicly funded research have become more common in the past few years. These include requirements of the National Institutes of Health [2], the Medical Research Council [3], and the Wellcome Trust [4]. Advocates of scientific data sharing such as the Science Commons network also strongly support this position:

'Research data, data sets, databases, and protocols should be in the public domain. This status ensures the ability to freely distribute, copy, re-format, and integrate data from research into new research, ensuring that as new technologies are developed researchers can apply those technologies without legal barriers. Scientific traditions of citation, attribution, and acknowledgment should be cultivated in norms' [5].

An article published in a meteorological journal describes data publication as an 'implicit part of the scientific method' [6], but very few clinical trialists currently make their raw data available. There are few strong incentives or requirements for doing so, nor is there a culture of data sharing, as has been established in other disciplines, such as the microarray [7] research community. Yet the benefits of sharing raw data have been recognised for many years.

RESEARCH METHODS & REPORTING

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

[1]BioMed Central, 236 Gray's Inn Road, London WC1X 8HL

[2]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, NY, NY 10021, USA

[3]Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Oxford OX2 6UD

Correspondence to: I Hrynaszkiewicz iain.hrynaszkiewicz@biomedcentral.com

Accepted: 11 December 2009

Many peer-reviewed journals' instructions for authors require that authors should be prepared to share their raw (that is, unprocessed) data with other scientists on request. Although data sharing is commonplace in some scientific disciplines and is a requirement of a number of major research funding agencies' policies, this culture has not yet been widely adopted by the clinical research community. Some journals have appealed to their authors to increase the availability of medical research data,[1-3] recognising the benefits of such transparency. These benefits are well documented and include replication of previous findings, comparisons with independent datasets, testing of additional hypotheses, teaching, and patient safety.[3-5] Moreover, patients themselves are increasingly seeing the benefits of openly sharing their experiences with others (www.patientslikeme.com/).

Online journals with unlimited space now provide the platform for publishing large, raw datasets as supplementary material,[6,7] but a common concern is confidentiality. If there is any doubt over anonymity, publishing data that have arisen from the doctor-patient or researcher-participant relationship will raise issues of privacy unless explicit and properly informed consent to all of the intended uses of that data has been obtained. The International Committee of Medical Journal Editors' *Uniform 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 95/46/EC) provides some harmony in data protection legislation, but in the US there is no overarching data protection law. Therefore, in an increasingly global research and publishing industry, universally agreed definitions as to what constitutes anonymised patient information would benefit clinical researchers. The HIPAA provides a list of 18 items that need to be removed from patient information in order for it to be considered anonymous for the purposes of sharing information between the "covered entities" specified in the act, but the list was not designed with publication in biomedical journals in mind. A number of publications from UK bodies provide some form of guidance on identifying information,[9-12] but none is as explicit as the HIPAA.

This article aims to provide practical guidance for those involved in the publication process by proposing a minimum standard for anonymising (or de-identifying) data for the purposes of publication in a peer reviewed biomedical journal or sharing with other researchers, either directly, where appropriate, or via a third party. Basic advice on file preparation is also provided, along with procedural guidance on prospective and retrospective publication of raw clinical data. Although the focus of this discussion is on data from randomised trials, the same issues of confidentiality apply to data from any research study involving human subjects, including cohort, case-control, and case series designs.

SUMMARY POINTS

- Despite journal and funder policies requiring data sharing, there has been little practical guidance on how data should be shared
- Confidentiality and anonymity are key considerations when publishing or sharing data relating to individuals, and this article provides practical advice on data sharing while minimising risks to patient privacy
- Consent for publication of appropriately anonymised raw data should ideally be sought from participants in clinical research
- Direct identifiers such as patients' names should be removed from datasets; datasets that contain those or more indirect identifiers, such as age or sex, should be reviewed by an independent researcher or ethics committee before being submitted for publication

Data preparation guidance

What is the dataset?

For the purposes of this guidance, the dataset is the aggregated collection of patient observations (including sociodemographic and clinical information) used for the purposes of producing the summary statistical findings presented in the main report of the research project, whether previously published or not.

Data are almost always collected at a greater level of detail than are reported in a journal article. For example, each participant in a pain study may complete a pain diary twice a day for 30 days, with the authors reporting "mean post-treatment pain" for one or more groups of partici-

Cite this as: *BMJ* 2010;340:c181 doi:10.1136/bmj.c181

Data publishing

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

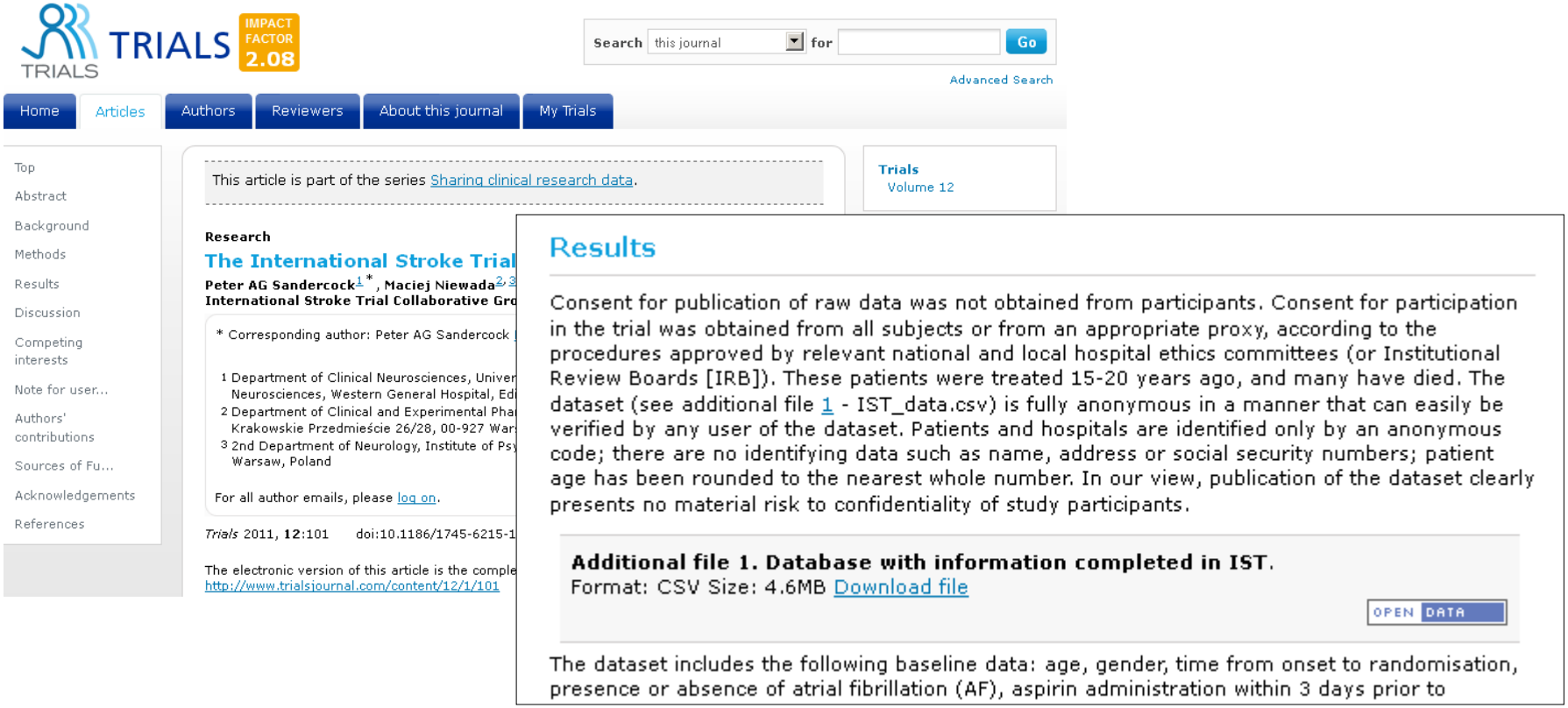
Associated material
PubMed record
About this article
Readers' comments
Pre-publication history

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

Sandercock *et al.*: The International Stroke Trial database. *Trials* 2011, **12**:101

BioMed Central
The Open Access Publisher

Data publishing



http://www.trialsjournal.com/series/sharing

BioMed Central
The Open Access Publisher

Data journals and repositories

http://www.gigasciencejournal.com

华大基因
BGI

DataCite

GigaScience is a new integrated database and journal co-published in collaboration between BGI Shenzhen and BioMed Central, to meet the needs of a new generation of biological and biomedical research as it enters the era of "big-data."

BioMed Central
The Open Access Publisher

Online enhancements

- Mini-websites as additional files at BioMed Central
- Embedded video (additional files)
- 3D structures (additional files)
- Online comments (rapid responses)
- Comments in context (e.g. PLoS Biology)
- Graphical abstracts (e.g. Chemistry Central)
- Animations (e.g. New England Journal of Medicine)
- 'Article of the future' (e.g. Cell Press, Elsevier)

Embedding multimedia

Ziegler *et al*. *BMC Medicine* 2011, 9:17
http://www.biomedcentral.com/1741-7015/9/17

COMMENTARY Open Access

Effectively incorporating selected multimedia content into medical publications

Alexander Ziegler[1*], Daniel Mietchen[2], Cornelius Faber[3], Wolfram von Hausen[4], Christoph Schöbel[5], Markus Sellerer[6], Andreas Ziegler[1]

Abstract

Until fairly recently, medical publications have been handicapped by being restricted to non-electronic formats, effectively preventing the dissemination of complex audiovisual and three-dimensional data. However, authors and readers could significantly profit from advances in electronic publishing that permit the inclusion of multimedia content directly into an article. For the first time, the *de facto* gold standard for scientific publishing, the portable document format (PDF), is used here as a platform to embed a video and an audio sequence of patient data into a publication. Fully interactive three-dimensional models of a face and a schematic representation of a human brain are also part of this publication. We discuss the potential of this approach and its impact on the communication of scientific medical data, particularly with regard to electronic and open access publications. Finally, we emphasise how medical teaching can benefit from this new tool and comment on the future of medical publishing.

Editorial note

During proofing and production of this article, there was significant debate about whether the multimedia files should be included as figures or additional files. We have taken the view that readers and indexers currently expect figures to be 2D graphical images suitable for printing, which is not the case with these files. The multimedia files are embedded into the PDF version of the article, and downloadable from links in the HTML version of the article. This article may act as a catalyst for Publishers to agree on the best way to present multimedia content.

Introduction

Despite substantial improvements since the early attempts of anatomists during the Renaissance, medical illustrations have always been handicapped by being restricted to two dimensions (2D). Comparable, but even more severe limitations have prevented the distribution of moving images as well as sounds through medical publications. A common solution to communicate complex multimedia data currently relies on the creation of supplemental files that are available for download either through the publisher's or the author's website. However, this results in the unattractive separation of the actual publication from supporting multimedia data, which may contain crucial information (see [1] for an example). As electronic publishing gains momentum, it seems logical to fully exploit its potential by integrating multimedia and text files into a single article.

While online publishing formats are gaining popularity, there can be no doubt that the current standard in electronic publishing is the portable document format (PDF). Since June 2008, this file type provides also the possibility to integrate three-dimensional (3D), video, and audio content together with the text into a single file. Strangely, the potential of this technique does not seem to have been recognised so far in the field of medical publishing, while astronomers [2], chemists [3], structural biologists [4,5], as well as zoologists [6,7] have already exploited the many opportunities offered by this approach. Using examples from various medical disciplines, we demonstrate how the readers of medical publications can profit from embedded multimedia content. We also point out some of the opportunities that are now available for medical publishing in general.

Interactive 3D imagery embedded into publications

To highlight the improvements that are attainable, we present here two fully interactive 3D models. The model of a face that is integrated into this article (Additional

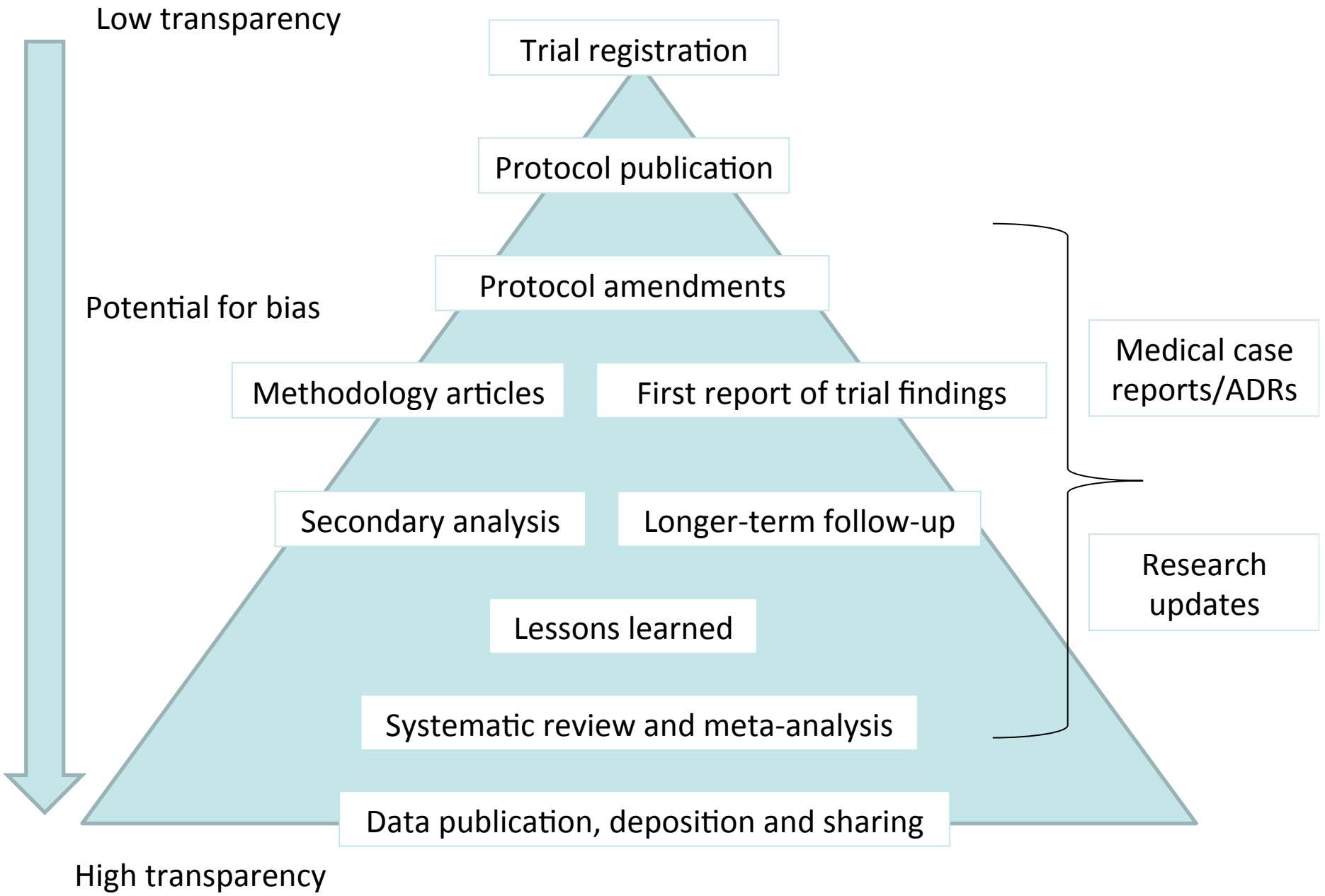
* Correspondence: alexander.ziegler@charite.de
[1]Institut für Immungenetik, Charité-Universitätsmedizin Berlin, Freie Universität Berlin, Berlin, Germany
Full list of author information is available at the end of the article

Beyond impact (factor)

- Citations (Google Scholar, Web of Science, Scopus, PubMed Central etc)
- Blog referrals/comments; article ratings
- Social media postings (e.g. Twitter, Facebook)
- Social bookmarking (e.g. Connotea)
- http://article-level-metrics.plos.org/
- http://total-impact.org/

Unlimited space

- Complete reporting of research (e.g. CONSORT)
- Avoidance of publication bias (scientifically sound small scale, confirmatory or negative results)
- Publication of datasets and data papers
- All trial-related publications (e.g. Registration records, study protocols, updates, case reports, methodology
- http://www.trialsjournal.com/series/5years


Low transparency
Potential for bias
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
Medical case reports/ADRs
Research updates
High transparency

‘Threaded publications’

In 1999 in *The Lancet*, Chalmers and Altman wrote:

“Electronic publication of a protocol could be simply the first element in a sequence of 'threaded' electronic publications, which continues with reports of the resulting research (published in sufficient detail to meet some of the criticisms of less detailed reports published in print journals), followed by deposition of the complete data set.”

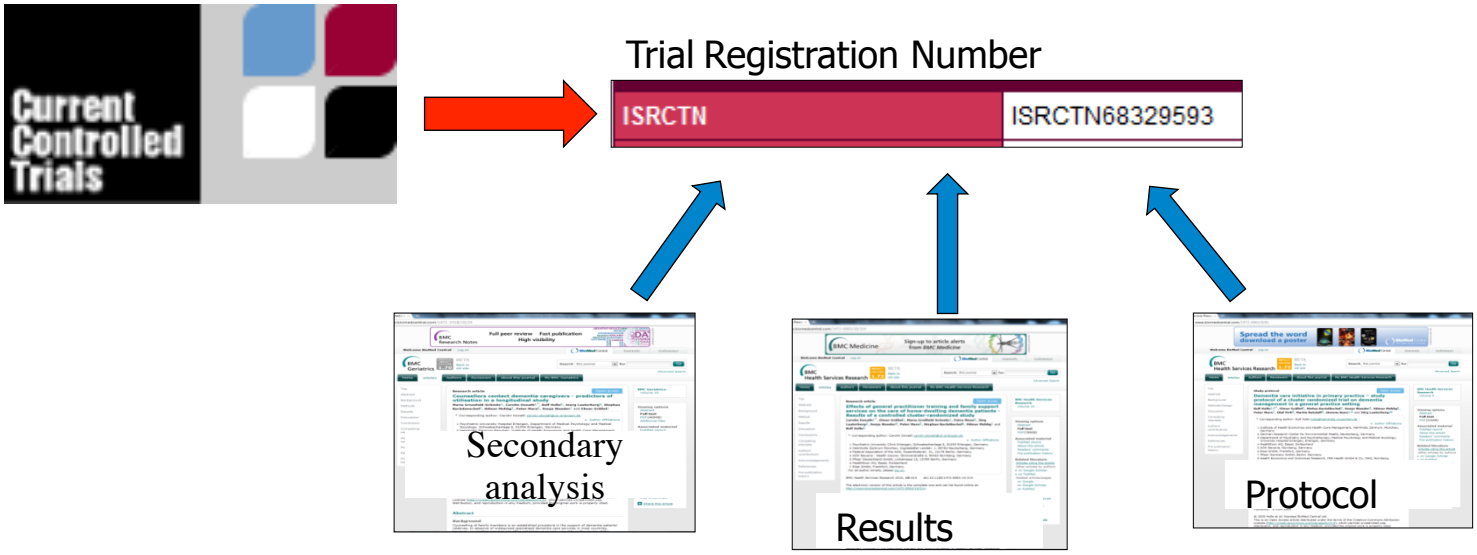
Chalmers I, Altman DG: **How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing.** *Lancet* 1999, **353:**490-493.

How does it work?

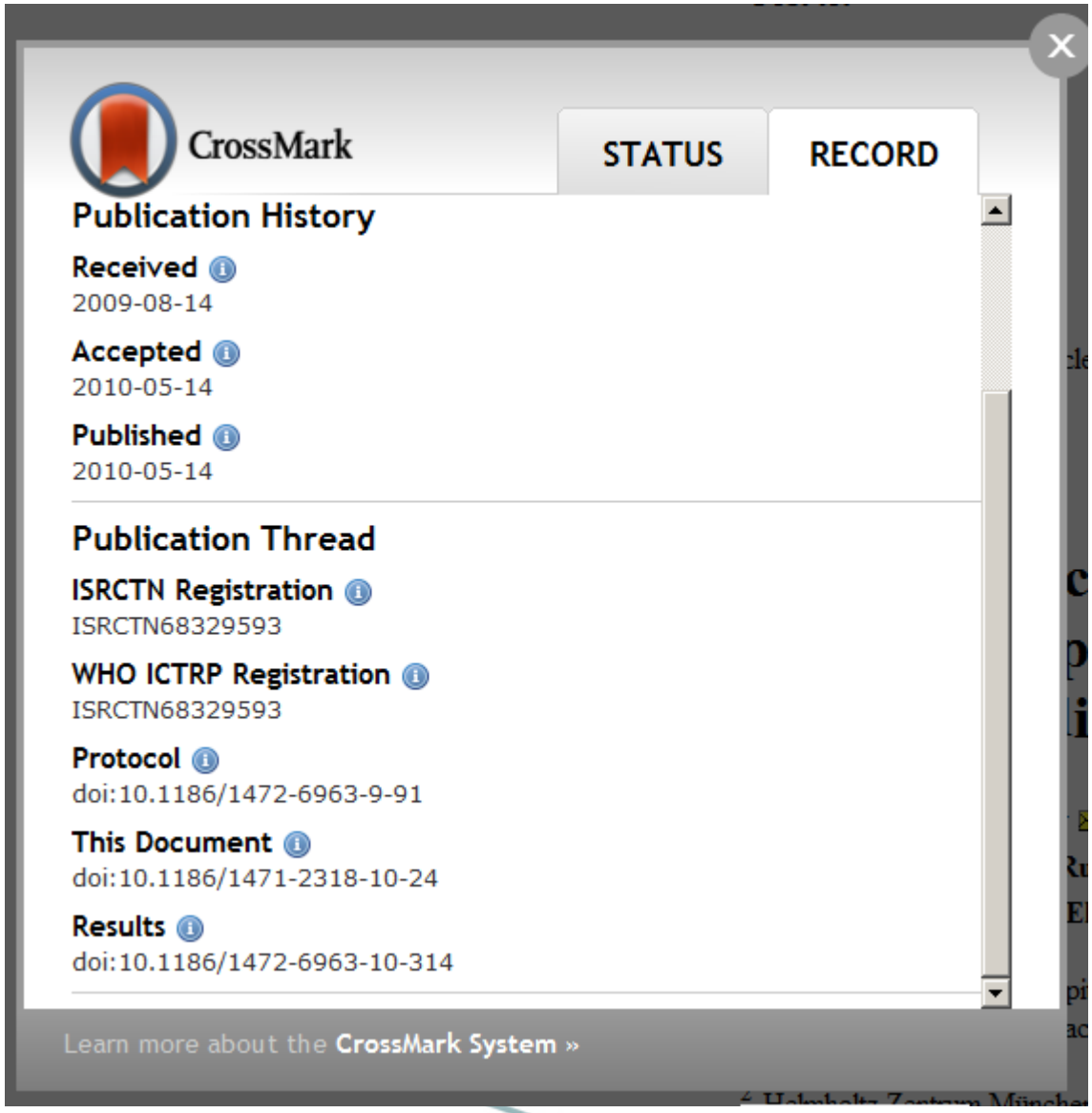
- Hyperlinks between trial registration records and trial protocol/results – or any other article including the trial ID in the abstract
- Article types and journals and editorial policies that enable publication of all clinical trial-related publications
- Financial – as well as the ethical and legal – incentives for authors, and funding agencies who ultimately often fund publication, to complete the scientific record

Beyond (hyper)linking

- No way to easily discern the relationships between related articles based on a common trial



- BMC articles hyper-link to major registries
- But links one-directional
- BMC working with CrossRef to develop CrossMark for threaded publications pan-publisher



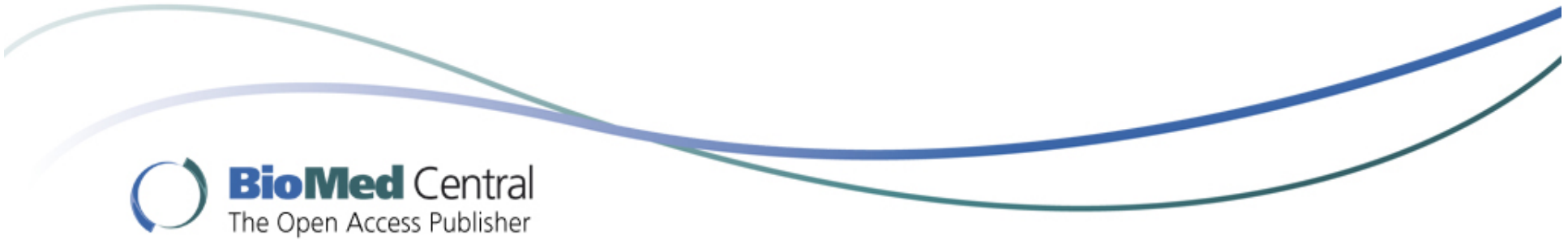
crossref

BioMed Central
The Open Access Publisher

"[P]ossibly the single most useful thing we could do to enhance the current literature" – Dr Cameron Neylon (Science and Technology Facilities Council)

BioMed Central
The Open Access Publisher

The next 10 years...



The next 10 years...

Information overload?

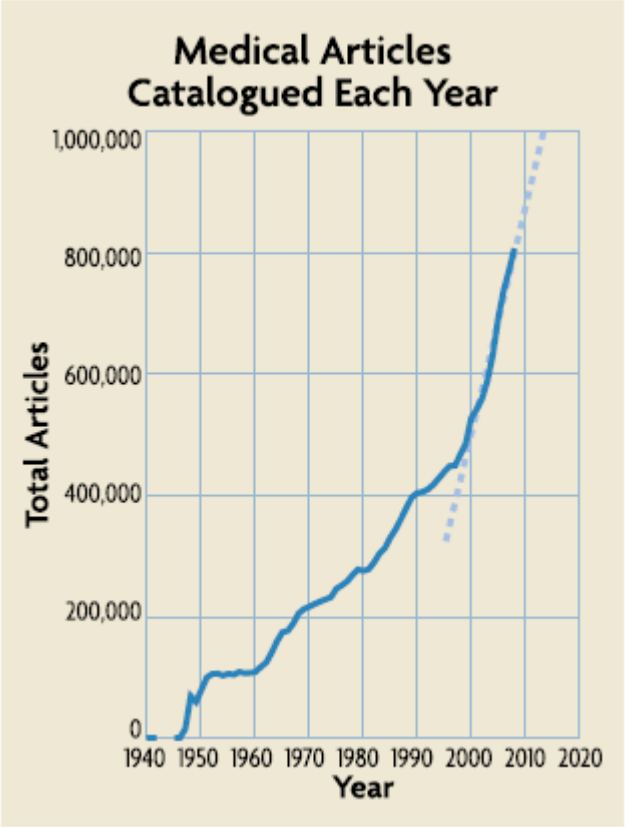


Image adapted from Gillam *et al*: **The Healthcare Singularity and the Age of Semantic Medicine**. In *The Fourth Paradigm (*2009)

BioMed Central
The Open Access Publisher

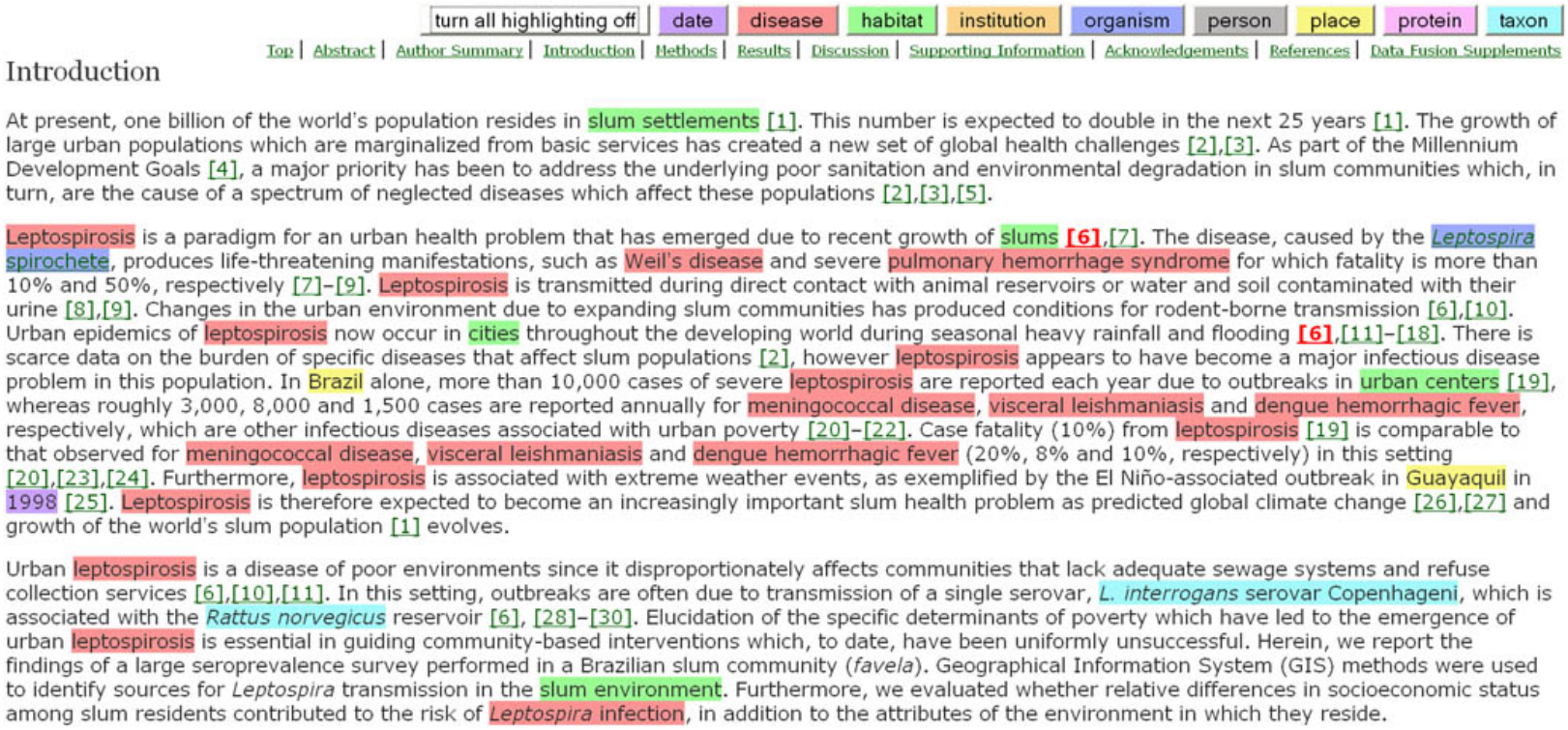
The next 10 years...

- More journals, formats, transparency
- Continued growth of open access
- Mobile applications and optimisation
- Data – journals, papers, visualisations, links
- Better connected literature
- More structured, customizable and interactive content
- More transparency, including study registration

The next 10 years...

- Diversification of impact measures
- Growth in emerging markets (e.g. China)
- Growing importance of post-publication peer review and 'social' services such as Mendeley, Papers, F1000 – to identify papers of interest and importance
- Emergence of secondary (e.g. semantically enhanced) products on OA literature

Adventures in semantics



turn all highlighting off | date | disease | habitat | institution | organism | person | place | protein | taxon

Top | Abstract | Author Summary | Introduction | Methods | Results | Discussion | Supporting Information | Acknowledgements | References | Data Fusion Supplements

Introduction

At present, one billion of the world's population resides in slum settlements [1]. This number is expected to double in the next 25 years [1]. The growth of large urban populations which are marginalized from basic services has created a new set of global health challenges [2],[3]. As part of the Millennium Development Goals [4], a major priority has been to address the underlying poor sanitation and environmental degradation in slum communities which, in turn, are the cause of a spectrum of neglected diseases which affect these populations [2],[3],[5].

Leptospirosis is a paradigm for an urban health problem that has emerged due to recent growth of slums [6],[7]. The disease, caused by the *Leptospira* spirochete, produces life-threatening manifestations, such as Weil's disease and severe pulmonary hemorrhage syndrome for which fatality is more than 10% and 50%, respectively [7]–[9]. Leptospirosis is transmitted during direct contact with animal reservoirs or water and soil contaminated with their urine [8],[9]. Changes in the urban environment due to expanding slum communities has produced conditions for rodent-borne transmission [6],[10]. Urban epidemics of leptospirosis now occur in cities throughout the developing world during seasonal heavy rainfall and flooding [6],[11]–[18]. There is scarce data on the burden of specific diseases that affect slum populations [2], however leptospirosis appears to have become a major infectious disease problem in this population. In Brazil alone, more than 10,000 cases of severe leptospirosis are reported each year due to outbreaks in urban centers [19], whereas roughly 3,000, 8,000 and 1,500 cases are reported annually for meningococcal disease, visceral leishmaniasis and dengue hemorrhagic fever, respectively, which are other infectious diseases associated with urban poverty [20]–[22]. Case fatality (10%) from leptospirosis [19] is comparable to that observed for meningococcal disease, visceral leishmaniasis and dengue hemorrhagic fever (20%, 8% and 10%, respectively) in this setting [20],[23],[24]. Furthermore, leptospirosis is associated with extreme weather events, as exemplified by the El Niño-associated outbreak in Guayaquil in 1998 [25]. Leptospirosis is therefore expected to become an increasingly important slum health problem as predicted global climate change [26],[27] and growth of the world's slum population [1] evolves.

Urban leptospirosis is a disease of poor environments since it disproportionately affects communities that lack adequate sewage systems and refuse collection services [6],[10],[11]. In this setting, outbreaks are often due to transmission of a single serovar, *L. interrogans* serovar Copenhageni, which is associated with the *Rattus norvegicus* reservoir [6], [28]–[30]. Elucidation of the specific determinants of poverty which have led to the emergence of urban leptospirosis is essential in guiding community-based interventions which, to date, have been uniformly unsuccessful. Herein, we report the findings of a large seroprevalence survey performed in a Brazilian slum community (*favela*). Geographical Information System (GIS) methods were used to identify sources for *Leptospira* transmission in the slum environment. Furthermore, we evaluated whether relative differences in socioeconomic status among slum residents contributed to the risk of *Leptospira* infection, in addition to the attributes of the environment in which they reside.

Shotton *et al*: PLoS Comput Biol 2009; 5(4): e1000361
http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1000361

Conclusions

- The internet has fundamentally changed publishing
- Open access and more open transferable peer review can enhance the published record, reduce bias, and increase efficiency and transparency
- The metrics of success for journals, articles, and authors are evolving
- Data and software are more integral to the scientific record but papers still help put data into context

Questions?

December 6th 2011

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

BioMed Central
The Open Access Publisher