Data without full methods: it's just a pile of … numbers

We support greater access to data, including public sector data and data generated from research. However, we assume that the panel recognizes that re-use of data requires much more than the numeric 'data'. To be usable and interpretable requires many other elements which are often either unavailable or poorly documented. The requirements might be expressed as a word equation of:

Usable Data = Full protocol + data dictionary + analysis plan + … + data

The lack of all these elements, or even summary data has led to considerable waste from Australia's, and the worlds, data production and research efforts.

We recently estimated that around 85% of current health and medical research is "wasted" – a global total of over $150 Billion annually[1]. The two major contributors to this waste are that about half of research is never published[2], and the published half is so poorly reported[3] that it is not replicable (by other researchers); not interpretable (because of selective or deceptive reporting); and not usable (by the end users of research). Hence even if "data" were publicly available, these problems from poor documentation and reporting would not be resolved. An international alliance to address these problems has been triggered by the Lancet series on research waste: http://rewardalliance.net/

Efforts to deposit, or make accessible, all data might paradoxically add to our current extraordinary waste, as researchers struggled with uninterpretable data or published erroneous results based on misinterpretations. Hence, we believe the problems of poor documentation and reporting must precede, or at least coincide with, greater public availability of study "data". This would include access to full protocols, data dictionaries, reproducible descriptions of measurements and interventions, etc. Without these the numbers are just a pile of … numbers.

Since a single positive result may arise from a lucky outcome, replication is a cornerstone of good science. For science to be replicable, a researcher needs to know the exact methods used in the design, conduct and analysis of the study. To interpret the results of a study, readers must be confident the results come from a reliable and well-designed, well-conducted, and correctly analysed study. That assessment requires complete reporting of all methods used in the study, a full description of the population studied, clear documentation of how many people stayed in the study to be measured with the outcomes of interest, and how those outcomes were measured. Groups have developed guidance for reporting of several study designs, but still authors have found that the standard of research papers has not improved dramatically. In 2007, the leading medical journals specified that the clinical trials they publish must have been registered in trials registries prior to commencement of patient recruitment. This was a valuable first step in reducing unnecessary duplication of research, and ensuring users of research could check that the published results were

produced using the methods specified in the register, and not altered post-hoc based on the study results. However, the registers currently cover only clinical trials, not other types of research, and the details in the register are too few to enable replicability or full usability.

The problem of documentation can be illustrated by one element of a clinical study - the treatment intervention. The intervention is the 'thing' that the researchers set out to evaluate - this might be a treatment (like a drug or exercise program), a public health policy, a mental health awareness program, or an education program delivered in schools. Without a full description of the intervention, effective interventions cannot reliably be used by others – this includes health professionals, patients, members of the public, policy makers, and anyone else with an interest in using the intervention. Other researchers, who might wish to replicate, or expand upon the research, are also without the details necessary for doing so.

There are now multiple studies, across many diverse fields of health research, highlighting the extent of this problem - with most finding over half of the studies examined had incomplete intervention descriptions[4], and hence were neither usable nor replicable by others. As intervention details are kept 'secret', others outside of the research team can, at best, only guess at what the intervention actually consisted of.

Much of this 'secret-keeping' appears to be inadvertent, with many researchers simply not aware of the importance of describing their interventions and what a full description actually consists of. Recognition of the importance of this aspect of research has only gained momentum in recent years. Recent initiatives to solve this problem have included producing a reporting statement (for researchers and journals) that provides guidance for how to fully describe an intervention. However systems and policies that facilitate (or mandate) this are also needed.

Waste in research can, and should, be decreased. Access to data, from public sector databases and from publicly funded research, are an important element, particularly as permanent data loss occurs rapidly -at about 7% per year[5]. However, without concurrent work on making the many other elements available, this may prove an expensive distraction or even counterproductive.

References

1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009 Jul 4;374(9683):86-9.

2. Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014 Jan 11;383(9912):101-4.

3. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014 Jan 18;383(9913):267-76.

4. Hoffmann TC, Erueti C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. BMJ. 2013 Sep 10;347:f3755.

5. Vines TH, Albert AY, Andrew RL et al. The availability of research data declines rapidly with article age. Curr Biol. 2014 Jan 6;24(1):94-7.

Submission Authors:

1. Professor Paul Glasziou, Director, Centre for Research in Evidence-Based Practice,

Faculty HS&M, Bond University, www.crebp.net.au

2. Professor Tammy Hoffmann, Centre for Research in Evidence-Based Practice,

Faculty HS&M, Bond University, www.crebp.net.au

3. A/Professor Elaine Beller, Australasian EQUATOR Centre, Faculty HS&M, Bond University, www.crebp.net.au