A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network

Douglas G Altman and Iveta Simera
Centre for Statistics in Medicine, University of Oxford, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, Botnar Research Centre, Oxford OX3 7LD, UK
Corresponding author: Doug Altman. Email: doug.altman@csm.ox.ac.uk

Introduction
Testing medical treatments and other interventions aimed at improving people’s health is incredibly important. However, comparative studies need to be well designed, well conducted, appropriately analysed and responsibly interpreted. Sadly, not all available findings and ‘discoveries’ are based on reliable research.

Our beliefs about best practices for medical research developed massively over the 20th century and ideas and methods continue to evolve. Much, perhaps most, medical research is done by individuals for whom it is not their main sphere of activity; notably, clinicians are expected to conduct some research early in their careers. As such, it is perhaps not surprising that there have been consistent comments on the poor quality of research and also recurrent attempts to raise understanding of how to do research well.

More recently, and increasingly over the last 20 years, concerns about poor methodology have been augmented by growing concerns about the inadequacy of reporting in published journal articles.5 His unfortunately brief summary of his observations about the 200 articles he examined was as follows:

In order to gain some knowledge of the degree to which statistical logic is being used, a survey was made of a sample of 200 medical-physiological quantitative papers from current American periodicals.5 His unfortunately brief summary of his observations about the 200 articles he examined was as follows:

- In over 90% statistical methods were necessary and not used.
- In about 85% considerable force could have been added to the argument if the probable error concept had been employed in one form or another.
- In almost 40% conclusions were made which could not have been proved without setting up some adequate statistical control.
- About half of the papers should never have been published as they stood; either because the numbers of observations were insufficient to prove the conclusions or because more statistical analysis was essential.

Statistical methods must eventually become an essential tool for the physiologist. It will be the physiologist who uses this tool most effectively and not the statistician untrained in physiological methods.

The earliest publication providing a detailed report of the weaknesses of a body of published research articles across specialties seems to have been that by Schor and Karten, respectively, a statistician and a medical student.11 They investigated the lack of statistical planning and evaluation in published articles and presented a programme to improve publications. They examined 295 publications in 10 of the ‘most frequently read’ medical journals between January and March 1964, of which 149 were analytical studies and...
146 case descriptions. Their main findings for the analytical studies were:

- 34% Conclusions drawn about population but no statistical tests applied on the sample to determine whether such conclusions were justified.
- 31% No use of statistical tests when needed.
- 25% Design of study not appropriate for solving problem stated.
- 19% Too much confidence placed on negative results with small-size samples.

Their bottom line summary was as follows: ‘Thus, in almost 73% of the reports read (those needing revision and those which should have been rejected), conclusions were drawn when the justification for these conclusions was invalid’.

Over the last 50 years occasional similar reviews have been published. It is common that the reviewers report that a high percentage of papers had methodological problems. A few examples are:

- In 100 orthopaedic research papers published in seven journals in 2005–2010, the conclusions were not clearly justified by the results in 17% and a different analysis should have been undertaken in 39%.
- Of 100 consecutive papers sent for review at the journal Injury, 47 used an inappropriate analysis.

In 4-yearly surveys of a sample of clinical trials reported in five general medical journals beginning in 1997, Clarke and his colleagues have drawn attention to the failure of authors and journals to ensure that the results of new trials include why the additional studies were done and what difference the results made to the accumulated evidence addressing the uncertainties in question.

**Reporting medical research**

The main focus of Schor and Karten was the use of valid methods and appropriate interpretation. Although they did not address reporting as such, any attempt to assess the appropriateness of methodology used in research runs quickly into the problem that the methods are often poorly described. For example, it is impossible to assess the extent to which bias was avoided without details of the method of allocation of trial participants to treatments. Likewise it is impossible to use the results of a trial in clinical practice if the article does not include full details of the interventions.
Concern about the completeness of research reports is a relatively recent phenomenon, strongly linked to the rise of systematic reviews. However, there are some early examples of the recognition of the importance of how research findings are communicated. The earliest such comments of which we are aware were made by the anatomist-turned-statistician Donald Mainland. In one of his earliest methodological publications, Mainland commented on the importance of how numerical results were presented, and he devoted a whole chapter of his 1938 textbook to ‘Publication of data and results’.

Assessing published reports of clinical trials
The earliest review we know devoted to assessment of published reports of clinical trials is that of Ross, who found that only 27 of 100 clinical trials were well controlled, and over half were uncontrolled.

Sandifer et al. studied 106 reports of trials of psychiatric drugs, aiming to compare those published before or after the report of Cole et al., which had included recommendations on reporting clinical trials. In so doing, they anticipated by some decades a before–after study design that has become quite familiar (Figure 1). Their detailed assessment included aspects of reporting of clinical details, interventions and methodology.

Another early study that looked at reporting, also in psychiatry, focused on whether authors gave adequate details of the interventions being tested. Glick wrote:

Two of the 29 studies did not indicate in any manner the duration of therapy. One of these was the paper which had given no dosage data. Thus 27 studies were left wherein there was some notation of duration. However, in four of these, duration was mentioned in such a vague or ambiguous way as to be unsuitable for comparative purposes. For instance, the duration of treatment might be given as ‘at least two months,’ or, ‘from one to several months.’

After these early studies there was a steady trickle of similar studies, examining the reporting of clinical trials in journal articles. Recent years have seen a vast number of such studies. Dechartres and colleagues identified 177 literature reviews published from 1987 to 2007, 58% of which were published...
after 2002. The rate has escalated further subsequently.

Developing the first reporting guidelines for randomised trials

The path to CONSORT

Many types of guidelines are relevant for clinical trials – they might relate to study conduct, reporting, critical appraisal or peer review. All of these could address the same important elements of trials, notably allocation to interventions, blinding, outcomes, interventions and completeness of data. These key elements also feature strongly in assessments of study ‘quality’ or ‘risk of bias’. However, an important criticism of many tools for assessing the quality of publications is that they mix considerations of methods (bias avoidance) with aspects of reporting.

Since the 1980s there had been occasional suggestions that it would be useful to have guidelines restricted to what should be reported (Box 3). Some of these authors suggested that medical journals should provide guidelines for authors.

There were occasional early calls for better reporting of randomised control trials (RCT) (Box 2), but the few early guidelines for reports of RCTs had very little impact. These guidelines tended to be targeted at reviewers.

A notable exception was the proposal that journal articles should have structured abstracts. First proposed in 1987 and updated in 1990, detailed guidelines were provided for abstracts of articles reporting original medical research or systematic reviews. Structured abstracts were quickly adopted by many medical journals, although they did not necessarily adhere to the detailed recommendations. There is now considerable evidence that structured abstracts communicate more effectively than traditional ones.

Serious attempts to develop guidelines relating to the reporting of complete research articles and targeted at authors began in the 1990s. In December 1994 two publications in leading general medical journals presented independently developed guidelines.

Figure 1. Some results from a 1961 review of published trials of psychiatric drugs by Sandifer et al.
Box 3. Early comments about the desirability of reporting guidelines.

- Standards governing the content and format of statistical aspects should be developed to guide authors in the preparation of manuscripts.46

- ... editors could greatly improve the reporting of clinical trials by providing authors with a list of items that they expected to be strictly reported.28

- An obvious proposal is to suggest that editors of oncology journals make up a check-list for authors of submitted clinical trial papers.37

- Unfortunately, in recent years I have become increasingly aware of the fact that it is very difficult to publish a manuscript which has been carefully written to communicate to the reader the key decisions that were made during the progress and analysis of the study, as there are many medical journal reviewers who consider these details irrelevant. The issue here is not whether the study was performed properly – it is whether it can be reported adequately. Clearly, there is a real need for further education of medical reviewers as to the information required to evaluate medical studies effectively. And there is a corresponding need for statisticians to develop reporting strategies more acceptable to the medical community than those currently available.38

- Authors should be provided with a list of items that are required. Existing checklists do not cover treatment allocation and baseline comparisons as comprehensively as we have suggested. Even if a checklist is given to authors there is no guarantee that all items will be dealt with. The same list can be used editorially, but this is time-consuming and inefficient. It would be better for authors to be required to complete a check list that indicates for each item the page and paragraph where the information is supplied. This would encourage better reporting and aid editorial assessment, thus raising the quality of published clinical trials.39

guidelines for reporting randomised controlled trials: ‘Asilomar’45 and SORT.46 Each had arisen from a meeting of researchers and editors concerned to try to improve the standard of reporting trials. Although the two checklists had overlapping content there were some notable differences in emphasis.

Of particular interest, the SORT recommendation arose from a meeting initially intended to develop a new scale to address the quality of RCT methodology, a key element of the conduct of a systematic reviews. Early in the meeting Tom Chalmers47 (JLL) argued that poor reporting of research was a major problem that undermined the assessment of published articles, so the meeting was redirected towards developing recommendations for reporting RCTs.46

The CONSORT Statement

Following the publication of the SORT and Asilomar recommendations, Drummond Rennie, Deputy Editor of JAMA, suggested that the SORT and Asilomar proposals should be combined into a single, coherent set of evidence-based recommendations.48 To that end, representatives from both groups met in Chicago in 1996 and produced the CONSolidated Standards Of Reporting Trials (CONSORT) Statement, published in 1996.49 The CONSORT Statement comprised a checklist and flow diagram for reporting the results of RCTs.

The rationale for including items in the checklist was that they are all necessary to evaluate a trial – readers need this information to be able to judge the reliability and relevance of the findings. Whenever possible, decisions to include items were based on relevant empirical evidence. The CONSORT recommendations were updated in 2001 and published simultaneously in three leading general medical journals.50 At the same time, a long 'Explanation and Elaboration' (E&E) paper was published, which included detailed explanations of the rationale for each of the checklist items, examples of good reporting and a summary of evidence about how well (or poorly) that information was reported in published reports of RCTs.51 Both the checklist and the E&E paper were updated again in 2010 in the light of new evidence.52,53

The checklist is seen as the minimum set of information, indicating information that is needed for all randomised trials. Clearly any important information about the trial should be reported, whether or not it is specifically addressed in the checklist. The flow diagram shows the passage of trial participants through the trial, from recruitment to final analysis (Figure 2). Although rare earlier examples exist,54 few reports of RCTs included flow diagrams prior to 1996. The flow diagram has become the most widely adopted of the CONSORT recommendations, although published diagrams often fail to include all the items recommended by CONSORT.55

The 2001 update of CONSORT clarified that the main focus was on two-arm parallel group trials.50 The first published CONSORT extensions addressed reporting of cluster randomised trials56 and
non-inferiority and equivalence trials. Both have been updated to take account of the changes in CONSORT 2010. A recent extension addressed N-of-1 trials. Design-specific extensions to the CONSORT checklist led to modification of some checklist items and the addition of some new elements to the checklist. Some also require modification of the flow diagram.
Two further extensions of CONSORT are relevant to almost all trial reports. They relate to the reporting of harms and the content of abstracts.

The influence of CONSORT on other reporting guidelines

The CONSORT Statement proved to be a very influential guideline that impacted not only on the way we report clinical trials but also on the development of many other reporting guidelines. Factors in the success of CONSORT include:

- Membership of CONSORT group includes methodologists, trialists and journal editors.
- Concentration on reporting rather than study conduct.
- Recommendations based on evidence where possible.
- Focus on the essential issues (i.e. the minimum set of information to report).
- High profile publications.
- Support from major editorial groups, hundreds of medical journals and some funding agencies.
- Dedicated executive group that coordinated ongoing developments and promotion.
- Updated to take account of new evidence and latest thinking.

The CONSORT approach has been adopted by several other groups. Indeed the QUOROM Statement (reporting recommendations for reporting meta-analyses of RCTs) was developed after a meeting held in October 1996, only a few months after the initial CONSORT Statement was published. The meeting to develop MOOSE (for reporting meta-analyses of observational studies) was held in April 1997. Several guidelines groups have followed CONSORT by producing detailed E&E papers to accompany a new reporting guideline, including STARD, STROBE, PRISMA, REMARK and TRIPOD.

CONSORT has also been the basis for guidelines for non-medical experimental studies, such as ARRIVE for in vivo experiments using animals, REFLECT for research on livestock and guidelines for software engineering.

The importance of implementation

Initial years of the EQUATOR Network

Reporting guidelines are important for achieving high standards in reporting health research studies. They specify minimum information needed for a complete and clear account of what was done and what was found during a particular kind of research study so the study can be fully understood, replicated, assessed and the findings used. Reporting guidelines focus on scientific content and thus complement journals’ instructions to authors, which mostly deal with the technicalities of submitted manuscripts. The primary role of reporting guidelines is to remind researchers what information to include in the manuscript, not to tell them how to do research.

In a similar way, they can be an efficient tool for peer reviewers to check the completeness of information in the manuscript. Judgements of completeness are not arbitrary: they relate closely to the reliability and usability of the findings presented in a report.

Potential users of research, for example systematic reviewers, clinical guideline developers, clinicians and sometimes patients, have to assess two key issues: the methodological soundness of the research (how well the study was designed and conducted) and its clinical relevance (how the study population relates to a specific population or patient, what the intervention was and how to use it successfully in practice, what the side effects encountered were, etc.). The key goal of a good reporting guideline is to help authors to ensure all necessary information is described sufficiently in a report of research.

Although CONSORT and other reporting guidelines started to influence the way research studies were reported, the documented improvement in adherence to these guidelines remained unacceptably low. To have a meaningful impact on preventing poor reporting, guidelines needed to be widely known and routinely used during the research publication process.

In 2006, one of us (DGA) obtained a one-year seed grant from the UK NHS National Knowledge Service (led by Muir Gray) to establish a programme to improve the quality of medical research reports available to UK clinicians through wider use of reporting guidelines. The initial project had three major objectives: (1) to map the current status of all activities aimed at preparing and disseminating guidelines on reporting health research studies; (2) to identify key individuals working in the area; and (3) to establish relationships with potential key stakeholders. We (DGA and IS) established a small working group with David Moher, Kenneth Schulz and John Hoey and laid the foundations of the new programme, which we named EQUATOR (Enhancing the QUAlity and Transparency Of health Research).

EQUATOR was the first coordinated attempt to tackle the problems of inadequate reporting systematically and on a global scale. The aim was to create an ‘umbrella’ organisation that would bring together researchers, medical journal editors, peer reviewers,
developers of reporting guidelines, research funding bodies and other collaborators with a mutual interest in improving the quality of research publications and of research itself. This philosophy led to the programme’s name change into ‘The EQUATOR Network’ (www.equator-network.org).

The EQUATOR Network held its first international working meeting in Oxford in May–June 2006. The 27 participants from 10 countries included representatives of reporting guideline development groups, journal editors, peer reviewers, medical writers and research funders. The objective of the meeting was to exchange experience in developing, using and implementing reporting guidelines and to outline priorities for future EQUATOR Network activities. Prior to that first EQUATOR meeting we had identified published reporting guidelines and had surveyed their authors to document how the guidelines had been developed and what problems had been encountered during their development.74 The survey results and meeting discussions helped us to prioritise the activities needed for a successful launch of the EQUATOR programme. These included the development of a centralised resource portal supporting good research reporting and a training programme, and support for the development of robust reporting guidelines.

The EQUATOR Network was officially launched at its inaugural meeting in London in June 2008. Since its launch, there have been a number of important milestones and a heartening impact on the promotion, uptake and development of reporting guidelines.

The EQUATOR Library for health research reporting is a free online resource that contains an extensive database of reporting guidelines and other resources supporting the responsible publication of health research. As of September 2015, 22,000 users access these resources every month and this number continues to grow. The production of new reporting guidelines has increased considerably in recent years. The EQUATOR database of reporting guidelines currently contains 282 guidelines (accessed on 22 September 2015). The backbone comprises about 10 core guidelines, each providing a generic reporting framework for a particular kind of study (e.g. CONSORT for randomised trials, STROBE for observational studies, PRISMA for systematic reviews, STARD for diagnostic test accuracy studies, TRIPOD for prognostic studies, CARE for case reports, ARRIVE for animal studies, etc.). Most guidelines, however, are targeted at specific clinical areas or aspects of research.

There are differences in the way individual guidelines were developed.74,75 At present the EQUATOR database is inclusive and does not apply any exclusion filter based on reporting guideline development methods. However, in order to ensure the robustness of guideline recommendations and their wide acceptability it is important that guidelines are developed in ways likely to be trustworthy. Based on experience gained in developing CONSORT and several other guidelines, the EQUATOR team published recommendations for future guideline developers76 and the Network supports developers in various ways.

Making all reporting guidelines known and easily available is the first step in their successful use. Promotion, education and training form another key part of the EQUATOR Network’s core programme. The EQUATOR team members give frequent presentations at meetings and conferences and organise workshops on the importance of good research reporting and reporting guidelines. EQUATOR courses target journal editors, peer reviewers, and most importantly, researchers – authors of scientific publications. Developing skills in early stage researchers is the key to a long-term change in research reporting standards. Journal editors play an important role too, not only as gatekeepers of good reporting quality but also in raising awareness of reporting shortcomings and directing authors to reliable reporting tools. A growing number of journals link to the EQUATOR resources and participate in and support EQUATOR activities.

Recent literature reviews have shown evidence of modest improvements in reporting over time for randomised trials (adherence to CONSORT)77 and diagnostic test accuracy studies (adherence to STARD).78 The present standards of reporting remain inadequate, however.

Further development of the EQUATOR Network

The EQUATOR programme is not a fixed term project but an ongoing programme of research support. The EQUATOR Network is gradually developing into a global initiative. Until 2014 most of the EQUATOR activities were carried out by the small core team based in Oxford, UK. In 2014 we launched three centres to expand EQUATOR activities: the UK EQUATOR Centre (also the EQUATOR Network’s head office), the French EQUATOR Centre and the Canadian EQUATOR Centre. The new centres will focus on national activities aimed at raising awareness and supporting adoption of good research reporting practices. The centres work with partner organisations and initiatives as well as contributing to the work of the EQUATOR Network as a whole.
The growing number of people involved in the EQUATOR work also fosters wider involvement in and reporting of ‘research on research’. Each centre has its own research programme relating to the overall goals of EQUATOR. Research topics include reviews of time trends in the nature and quality of publications; the development of tools and strategies to improve the planning, design, conduct, management and reporting of biomedical research; investigating strategies to help journals to improve the quality of manuscripts; and so on (e.g. Hopewell et al., Stevens et al., Barnes et al., Mahady et al.).

In conclusion

Concern about the quality of medical research has been expressed intermittently over a century and quality about reports of research for almost as long. At last, in the 1990s, serious international efforts began to promote better reporting of medical research. The emergence of the EQUATOR Network has been both a result and a cause of the progress that is being made.

Declarations

Competing interests: None declared

Funding: The EQUATOR Network is supported by the UK NHS National Institute for Health Research, UK Medical Research Council, Cancer Research UK and the Pan American Health Organization. We would like to thank to these funding bodies

Acknowledgement:

None.


Greenwood M. What is wrong with the medical curriculum. Lancet 1932; i: 1269–1270.


34. Dechartres A, Charles P, Hopewell S, Ravaud P and Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol* 2011; 64: 136–144.


