

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 methodology¹ have been augmented by growing concerns about the inadequacy of reporting in published journal articles.²

Evaluating the quality of medical research

From at least as early as the first part of the 20th century, there have been publications referring disparagingly to the quality of research methods and inadequate understanding of research methodology, as judged by comparisons with prevailing standards (Box 1).

Halbert Dunn was a medical doctor subsequently employed as a statistician at the Mayo Clinic. He is probably best known for introducing, many years later, the concept of 'wellness'.¹⁰ Dunn may have been the first person to publish the findings of

a review of an explicit sample of journal publications.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. Here is the result:

- 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

Box 1. Early comments about poor research methodology.

'The quality of the published papers is a fair reflection of the deficiencies of what is still the common type of clinical evidence. A little thought suffices to show that the greater part cannot be taken as serious evidence at all.'³

'It is a commonplace of medical literature to find the most positive and sweeping conclusions drawn from a sample so meager as to make scientifically sound conclusions of any sort utterly impossible.'⁴

'Statistical workers who fail to scrutinize the goodness of their observed data and carry through a satisfactory analysis upon poor observations, will end up with ridiculous conclusions which cannot be maintained.'⁵

'Medical papers now frequently contain statistical analyses, and sometimes these analyses are correct, but the writers violate quite as often as before, the fundamental principles of statistical or of general logical reasoning . . . The writer, who 20 years ago would have said that statistical method was mere "mathematical" juggling having no relation to "practical" matters, now seeks for some technical "formula" by the application of which he can produce "significant" results . . . the change has been from thinking badly to not thinking at all.'⁶

'My own survey was not numerical and was concerned more with clinical than with laboratory medicine, but it revealed that the same general verdict, perhaps even a more adverse one, was appropriate in the clinical field. . . Frequently, indeed, the way in which the observations were planned must have made it impossible for the observer to form a valid estimate of the error . . . an idea of what results might be expected if the experiment were repeated under the same conditions.'⁷

' . . . less than 1% of research workers clearly apprehend the rationale of the statistical techniques they commonly invoke.'⁸

' . . . almost any volume of a medical journal contains faults that can be detected by first-year students after only three or four hours' guidance in the scrutiny of reports.'⁹

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:

- Among 513 behavioural, systems and cognitive neuroscience articles published in five top-ranking journals (*Science*, *Nature*, *Nature Neuroscience*, *Neuron* and *The Journal of Neuroscience*) in 2009–10, 50% of 157 articles which compared effect sizes used an inappropriate method of analysis.¹²
- 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%.¹³

- Of 100 consecutive papers sent for review at the journal *Injury*, 47 used an inappropriate analysis.¹⁴

In 4-yearly surveys of a sample of clinical trials reported in five general medical journals beginning in 1997,¹⁵ 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.¹⁶

Reporting medical research

The main focus of Schor and Karten¹¹ 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.¹⁷

Box 2. Early comments about reporting research.

'The way to a more adequate understanding and treatment of medical data would be opened up if all records, articles, and even abstracts gave, besides averages, the numbers of observations and the variation, properly expressed, e.g. as standard deviation (maxima and minima being very unreliable).'¹⁸

'... incompleteness of evidence is not merely a failure to satisfy a few highly critical readers. It not infrequently makes the data that are presented of little or no value.'⁷

'This leads one to consider if it is possible, in planning a trial, in reporting the results, or in assessing the published reports of trials, to apply criteria which must be satisfied if the analysis is to be entirely acceptable. . . . A high standard must be set, however, not only in order to assess the validity of results, but also because pioneering investigations of this type may in many ways serve as a model and lesson to future investigators. A basic principle can be set up that, just as in a laboratory experiment report, it is at least as important to describe the techniques employed and the conditions in which the experiment was conducted, as to give the detailed statistical analysis of results.'¹⁹

'A clinical experiment is not completed until the results have been made available to one's colleagues and co-workers. There is, in a sense, a moral obligation to "give posterity" the fruits of one's scientific labor. Certainly it would be a sad waste of effort to allow reams of data to lie yellowing in a dusty file, while in other laboratories workers are unnecessarily repeating the study.'²⁰

'Words like "random assignment", "blindfold technique", "objective methods" and "statistical analysis," are no guarantee of quality. The reader should ask: "What is the evidence that the investigator was keenly aware of what might interfere with the effects of the randomization, such as leakage in the blindfold, and pseudo-objective assessments?" "What steps did he take to prevent such risks?"'²¹

'It is difficult enough for a clinician to interpret the statistical meaning of a procedure with which he is unfamiliar; it is much more difficult when he is not told what that procedure was.'²²

'... the idea is to give all of the information to help others to judge the value of your contribution; not just the information that leads to judgement in one particular direction or another.'²³

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,¹⁸ and he devoted a whole chapter of his 1938 textbook to 'Publication of data and results'⁷ (Box 2).

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.²⁴

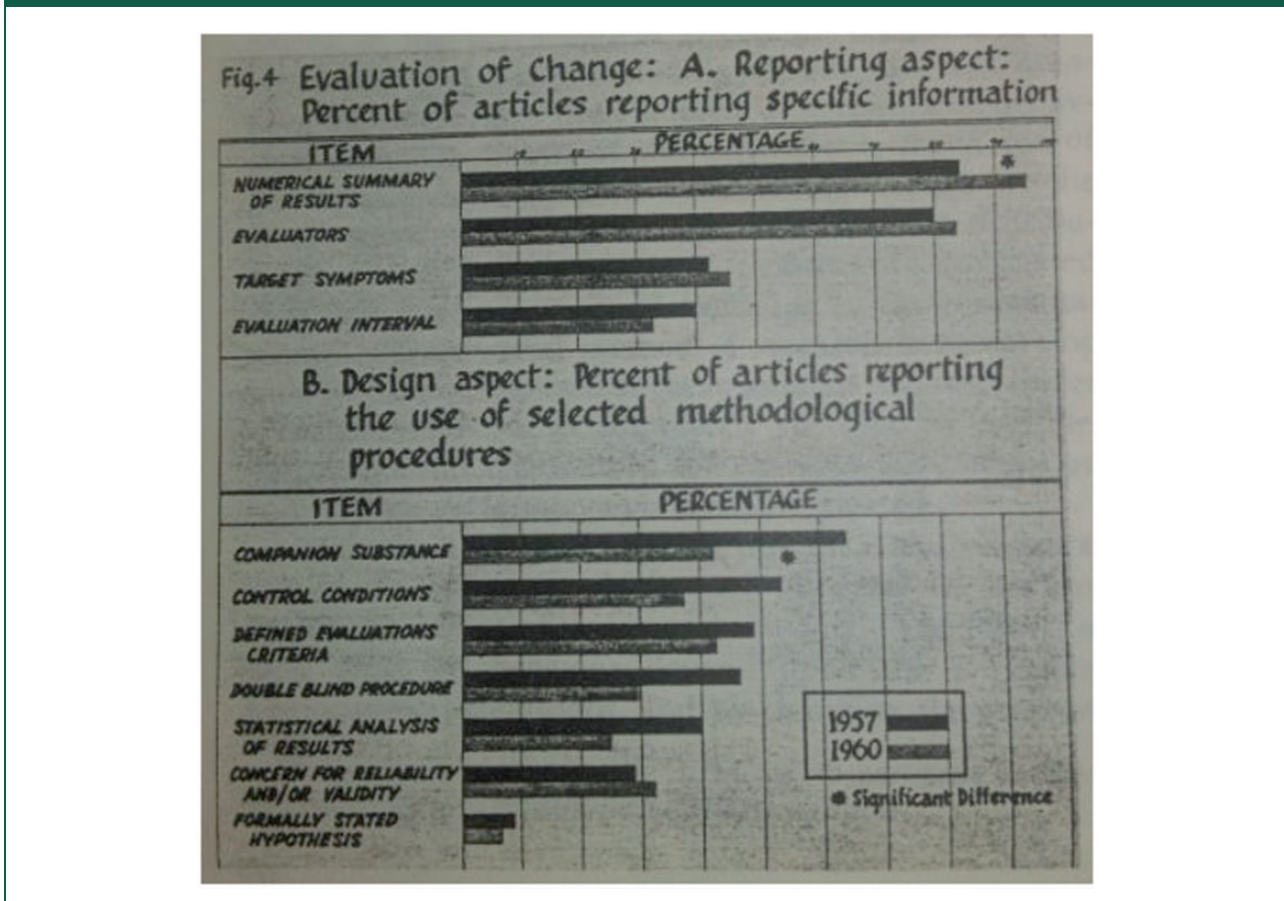
Sandifer et al.²⁵ studied 106 reports of trials of psychiatric drugs, aiming to compare those published before or after the report of Cole et al., which had included recommendations on reporting clinical trials.²⁶ In so doing, they anticipated by

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".²⁷

After these early studies there was a steady trickle of similar studies, examining the reporting of clinical trials in journal articles.^{28–33} 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

Figure 1. Some results from a 1961 review of published trials of psychiatric drugs by Sandifer et al.²⁵

after 2002.³⁴ 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.³⁵

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^{40,41} 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 1987⁴² and updated in 1990,⁴³ 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.⁴⁴

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

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.'³⁶

'... 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

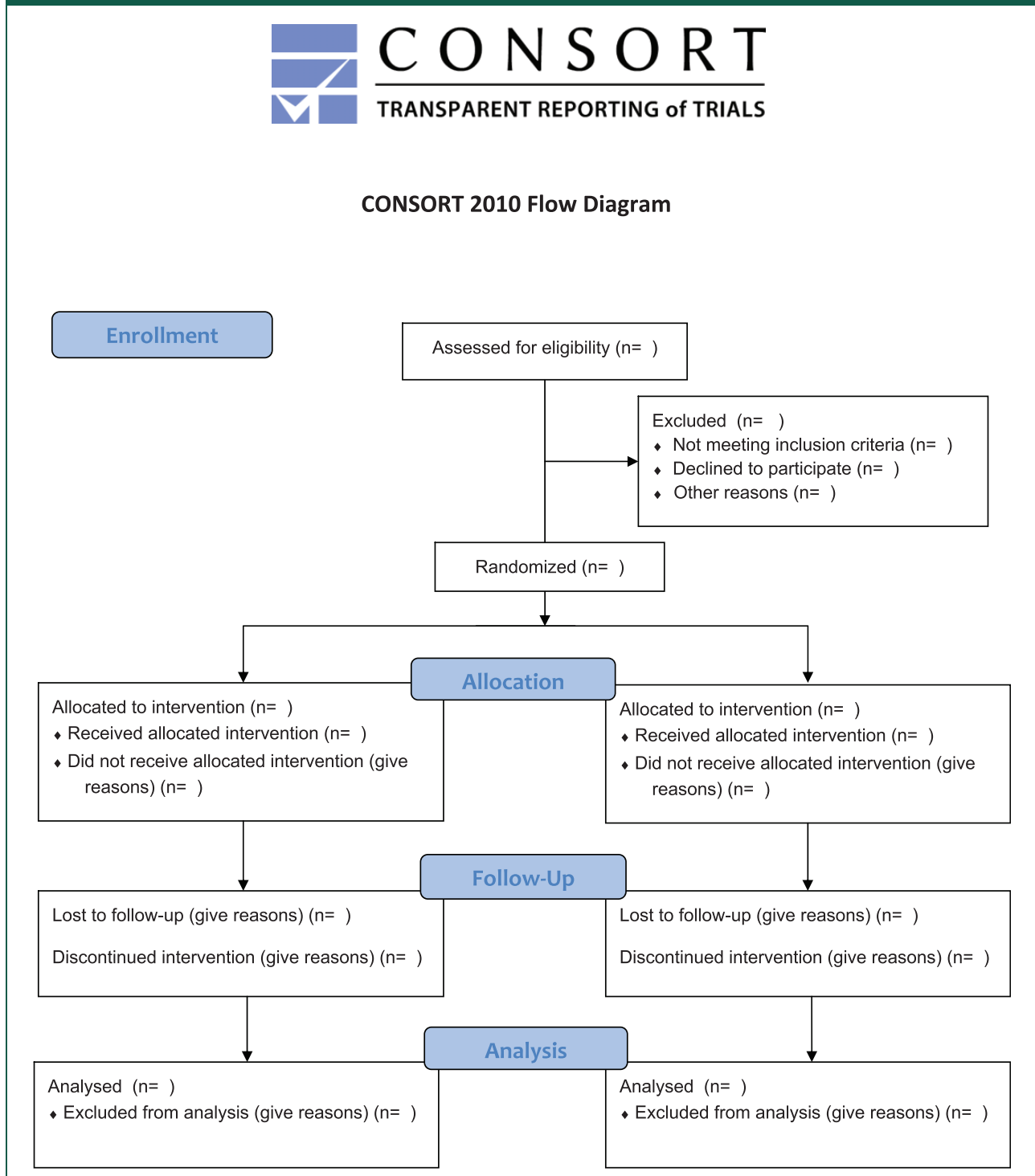
Following the publication of the SORT and Asilomar recommendations, Drummond Rennie, Deputy Editor of *JAMA*, suggested that the SORT and Asilomar proposals should be combined into a single, coherent set of evidence-based recommendations.⁴⁸ To that end, representatives from both groups met in Chicago 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

Figure 2. The CONSORT 2010 flow diagram.⁵³

non-inferiority and equivalence trials.⁵⁷ Both have been updated to take account of the changes in CONSORT 2010. A recent extension addressed N-of-1 trials.⁵⁸ 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 harms⁵⁹ and the content of abstracts.^{60,61}

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 1996,⁶² 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 1997.⁶³ Several guidelines groups have followed CONSORT by producing detailed E&E papers to accompany a new reporting guideline, including STARD,⁶⁴ STROBE,⁶⁵ PRISMA,⁶⁶ REMARK⁶⁷ and TRIPOD.⁶⁸

CONSORT has also been the basis for guidelines for non-medical experimental studies, such as ARRIVE for *in vivo* experiments using animals,⁶⁹ REFLECT for research on livestock⁷⁰ and guidelines for software engineering.⁷¹

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.^{72,73} 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.,⁷⁹ Stevens et al.,⁸⁰ Barnes et al.,⁸¹ Mahady et al.⁸²).

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 for providing vital financial support without which this programme would not be able to start and continue its development. None of the above funders influenced in any way the content of this article.

Ethical approval: Not applicable.

Guarantor: DGA.

Contributorship: DGA and IS jointly wrote and agreed the text.

Provenance: Invited contribution from the James Lind library

Acknowledgements: None.

References

- Altman DG. The scandal of poor medical research. *BMJ* 1994; 308: 283–284.
- Simera I and Altman DG. Writing a research article that is “fit for purpose”: EQUATOR Network and reporting guidelines. *Evidence-Based Med* 2009; 14: 132–134.
- Sollmann T. The crucial test of therapeutic evidence. *JAMA* 1917; 69: 198–199.
- Pearl R. A statistical discussion of the relative efficacy of different methods of treating pneumonia. *Arch Intern Med* 1919; 24: 398–403.
- Dunn HL. Application of statistical methods in physiology. *Physiol Rev* 1929; 9: 275–398.
- Greenwood M. What is wrong with the medical curriculum. *Lancet* 1932; i: 1269–1270.
- Mainland D. *The Treatment of Clinical and Laboratory Data*. Edinburgh: Oliver & Boyd, 1938.
- Hogben L. *Chance and Choice by Cardpack and Chessboard*. Vol. 1, New York, NY: Chanticleer Press, 1950.
- Mainland D. *Elementary Medical Statistics. The Principles of Quantitative Medicine*. Philadelphia, PA: WB Saunders, 1952.
- Dunn HL. What high-level wellness means. *Can J Public Health* 1959; 50: 447–457.
- Schor S and Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966; 195: 1123–1128.
- Nieuwenhuis S, Forstmann BU and Wagenmakers EJ. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci* 2011; 14: 1105–1107.
- Parsons NR, Price CL, Hiskens R, Achten J and Costa ML. An evaluation of the quality of statistical design and analysis of published medical research: results from a systematic survey of general orthopaedic journals. *BMC Med Res Methodol* 2012; 12: 60.
- Prescott RJ and Civil I. Lies, damn lies and statistics: errors and omission in papers submitted to *INJURY* 2010–2012. *Injury* 2013; 44: 6–11.
- Clarke M and Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA* 1998; 280: 280–282.
- Clarke M and Hopewell S. Many reports of randomised trials still don’t begin or end with a systematic review of the relevant evidence. *J Bahrain Med Soc* 2013; 24: 145–147.
- Glasziou P, Meats E, Heneghan C and Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008; 336: 1472–1474.
- Mainland D. Chance and the blood count. *Can Med Assoc J* 1934; 30: 656–658.
- Daniels M. Scientific appraisal of new drugs in tuberculosis. *Am Rev Tuberc* 1950; 61: 751–756.
- Waife SO. Problems of publication. In: Waife SO and Shapiro AP (eds) *The Clinical Evaluation of New Drugs*. New York, NY: Hoeber-Harper, 1959, pp.213–216.
- Mainland D. Some research terms for beginners: definitions, comments, and examples. I. *Clin Pharmacol Ther* 1969; 10: 714–736.
- Feinstein AR. Clinical biostatistics. XXV. A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 1974; 15: 97–107.
- Feynman R. Cargo cult science. *Eng Sci* 1974; 37: 10–13.
- Ross OB Jr. Use of controls in medical research. *JAMA* 1951; 145: 72–75.
- Sandifer MG, Dunham RM and Howard K. The reporting and design of research on psychiatric drug treatment: a comparison of two years. *Psychopharmacol Serv Cent Bull* 1961; 1: 6–10.
- Cole JO, Ross S and Bouthilet L. Recommendations for reporting studies of psychiatric drugs. *Public Health Rep* 1957; 72: 638–645.
- Glick BS. Inadequacies in the reporting of clinical drug research. *Psychiatr Q* 1963; 37: 234–244.

28. DerSimonian R, Charette LJ, McPeck B and Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1332–1337.
29. Tyson JE, Furzan JA, Reisch JS and Mize SG. An evaluation of the quality of therapeutic studies in perinatal medicine. *J Pediatr* 1983; 102: 10–13.
30. Meinert CL, Tonascia S and Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984; 5: 328–347.
31. Liberati A, Himel HN and Chalmers TC. A quality assessment of randomised controlled trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4: 942–951.
32. Pocock SJ, Hughes MD and Lee RJ. Statistical problems in the reporting of clinical trials. *NEJM* 1987; 317: 426–432.
33. Gøtzsche P. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10: 31–56.
34. Dechartres A, Charles P, Hopewell S, Ravaut 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.
35. Jüni P, Altman DG and Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42–46.
36. O'Fallon JR, Dubey SD, Salsburg DS, Edmonson JH, Soffer A and Colton T. Should there be statistical guidelines for medical research papers? *Biometrics* 1978; 34: 687–695.
37. Zelen M. The reporting of clinical trials: counting is not easy. *J Clin Oncol* 1989; 7: 827–828.
38. O'Fallon JR. Discussion of: Ellenberg JH. Biostatistical collaboration in medical research. *Biometrics* 1990; 45: 24–26.
39. Altman DG and Doré C. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149–153.
40. Grant A. Reporting controlled trials. *Br J Obstet Gynaecol* 1989; 96: 397–400.
41. Squires BP and Elmslie TJ. Reports of randomized controlled trials: what editors want from authors and peer reviewers. *CMAJ* 1990; 143: 381–382.
42. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598–604.
43. Haynes RB, Mulrow CD, Huth EJ, Altman DG and Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113: 69–76.
44. Hartley J. Current findings from research on structured abstracts: an update. *J Med Libr Assoc* 2014; 102: 146–148.
45. Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med* 1994; 121: 894–895.
46. Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272: 1926–1931.
47. Dickersin K and Chalmers F. Thomas C Chalmers (1917–1995): *A Pioneer of Randomized Clinical Trials and Systematic Reviews*. *JLL Bulletin: Commentaries on the History of Treatment Evaluation*. See <http://www.jameslindlibrary.org/articles/thomas-c-chalmers-1917-1995/>, 2014 (last checked 7 January 2016).
48. Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *JAMA* 1995; 273: 1054–1055.
49. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276: 637–639.
50. Moher D, Schulz KF and Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357: 1191–1194.
51. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663–694.
52. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
53. Schulz KF, Altman DG and Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.
54. Yelle L, Bergsagel D, Basco V, Brown T, Bush R, Gillies J, et al. Combined modality therapy of Hodgkin's disease: 10-year results of National Cancer Institute of Canada Clinical Trials Group multicenter clinical trial. *J Clin Oncol* 1991; 9: 1983–1993.
55. Hopewell S, Hirst A, Collins GS, Mallett S, Yu LM and Altman DG. Reporting of participant flow diagrams in published reports of randomized trials. *Trials* 2011; 12: 253.
56. Campbell MK, Piaggio G, Elbourne DR and Altman DG for the CONSORT Group. CONSORT 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345: e5661.
57. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ and Altman DG for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012; 308: 2594–2604.
58. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; 350: h1738.
59. Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz KF, et al. Improving the reporting of harms in randomized trials: expansion of the

- CONSORT statement. *Ann Intern Med* 2004; 141: 781–788.
60. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008; 371: 281–283.
 61. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008; 5: e20.
 62. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D and Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; 354: 1896–1900.
 63. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283: 2008–2012.
 64. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1–12.
 65. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4: e297.
 66. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
 67. Altman DG, McShane LM, Sauerbrei W and Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012; 10: 51.
 68. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1–73.
 69. Kilkenny C, Browne WJ, Cuthill IC, Emerson M and Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010; 8: e1000412.
 70. Sargeant JM, O'Connor AM, Gardner IA, Dickson JS and Torrence ME. The REFLECT statement: reporting guidelines for Randomized Controlled Trials in livestock and food safety: explanation and elaboration. *Zoonoses Public Health* 2010; 57: 105–136.
 71. Jedlitschka A, Pfahl D. Reporting guidelines for controlled experiments in software engineering. IEEE Computer Society. 4th International Symposium on Empirical Software Engineering (ISESE), Proceedings, 2005: 95–104.
 72. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329: 883.
 73. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006; 185: 263–267.
 74. Simera I, Altman DG, Moher D, Schulz KF and Hoey J. Guidelines for reporting health research: the EQUATOR Network's survey of guideline authors. *PLoS Med* 2008; 5: e139.
 75. Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol* 2011; 64: 718–742.
 76. Moher D, Schulz KF, Simera I and Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010; 7: e1000217.
 77. Turner L, Shamseer L, Altman DG, Schulz KF and Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012; 1: 60.
 78. Korevaar DA, Wang J, van Enst WA, Leeftang MM, Hooft L, Smidt N, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology* 2015; 274: 781–789.
 79. Hopewell S, Collins GS, Boutron I, Yu LM, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014; 349: g4145.
 80. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ* 2014; 348: g3804.
 81. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG and Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC Med* 2015; 13: 221.
 82. Mahady SE, Schlub T, Bero L, Moher D, Tovey D, George J, et al. Side effects are incompletely reported among systematic reviews in gastroenterology. *J Clin Epidemiol* 2015; 68: 144–153.