

## **The Availability of Source Data and Statistics: An Improvement in Good Publication Practice**

**Antonio M Morselli-Labate**

Department of Internal Medicine and Gastroenterology, Alma Mater Studiorum - University of Bologna. Bologna, Italy

### **Summary**

The purpose of this paper is to highlight the aspects of good publication practices, with particular reference to data analysis, and to propose an innovative initiative for improving the quality of scientific information in this field.

Several committees within the scientific community provide information and publish guidelines in order to support scientists in the application of good publication practices and to improve quality in medical research. Those guidelines suggest that the possibility of verifying the source data warrants the reliability of the published results by reducing the occurrence of misconduct related to data analysis.

The initiative proposed in this article is aimed at making the source data and the statistical reports available to the scientific community together with the actual paper. Such a practice is undoubtedly an improvement in the quality of publication permitting verification of the results as well as allowing for further elaboration of the same data.

effects of advancement in the many diverse scientific fields, (from biotechnology to communication, from technology to computer science, etc.) are extremely evident. All this is especially true in clinical studies which have an immediate impact on every day clinical practice. Just to give an example, I will mention the case of simvastatin in the prevention of acute myocardial infarction [1, 2].

Experimental study quality, in terms of carrying out and the publication of the results, is crucial to correct scientific information and its importance constantly increases within the scientific community, particularly within the biomedical one. Not to apply, or to incorrectly apply good publication practice criteria, leads to misconduct whose primary effect can be summarized as “causing others to regard as true that which is not true”. The impact of this statement in the scientific field is devastating. The following considerations are aimed at highlighting the aspects of good publication practice, with particular reference to data analysis, and at proposing an innovative initiative in order to improve the quality of scientific information in this field.

---

### **Introduction**

Nowadays, scientific information plays a fundamental role, considering that the knowledge and the application of the results obtained in the scientific field have very important consequences for our society. It is superfluous to give examples, since the

### **Good Publication Practice**

In the last few years, several committees were founded within the scientific community for the purpose of dealing with the problem of quality in scientific communication. Among these are COPE: Committee on Publication Ethics (<http://www.publicationethics.org.uk>);

CSE: Council of Science Editors (<http://www.councilscienceeditors.org>), formerly the CBE: Council of Biology Editors (<http://www.cbe.org>); EASE: European Association of Science Editors ([www.ease.org.uk](http://www.ease.org.uk)); SSP: Society for Scholarly Publishing ([www.sspnet.org](http://www.sspnet.org)) and WAME: Word Association of Medical Editors (<http://www.wame.org>). At the same time, other committees were funded with the specific objective of dealing with the aspects of quality in medical research. Among them are: ASSERT: A Standard for the Scientific and Ethical Review of Trials (<http://www.assert-statement.org>); CHA: Center for Health Affairs of the Project Health Opportunities for People Everywhere (HOPE) (<http://www.projecthope.org/CHA>); CIOMS: Council for International Organizations of Medical Sciences (<http://www.cioms.ch>); CONSORT: Consolidated Standards of Reporting Trials (<http://www.consort-statement.org>); ICH: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (<http://www.ich.org>); ORI: US Office on Research Integrity (<http://ori.dhhs.gov>).

One of the tasks of these committees is to provide information and publish guidelines in order to support scientists in the application of good publication practices. There are many aspects that must be taken into account in relation to the good quality of scientific information. Those aspects refer to topics which involve many different disciplines, as well as people with different roles. Among them, authors and publishers are certainly the most important ones, while other entities can be interested, even if not directly involved, in the scientific publication process. Some aspects, such as peer-reviewing, specifically concern the editors and the authors themselves, in that such activity is done by people who produce scientific information - therefore authors - in their turn. Other aspects more specifically interest the authors and are related to problems of various kinds: i.e. methodological problems (correctness of the experimental design, data analysis, etc.),

ethical problems, etc. There are also aspects that may concern all those involved in the publication process: i.e. conflicts of interest, plagiarism, redundant publications. Finally, they may also concern people not directly involved in the publication process, such as journalists (i.e., media relations). All those aspects have been extensively analyzed and discussed for many years now, and several reports and guidelines can be found at the Committees' websites (see above). COPE guidelines [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20] and summaries of the WAME report [21, 22, 23, 24, 25, 26] are also available in biomedical literature.

As to what more specifically concerns the preparation of manuscripts submitted to biomedical journals, the problem was initially considered by a small group of editors of general medical journals that met informally in Vancouver, British Columbia in 1978 and established guidelines for the format of manuscripts submitted to their journals. The group became known as the Vancouver Group. Its requirements for manuscripts were first published in 1979. The Vancouver Group expanded and evolved into the International Committee of Medical Journal Editors (ICMJE), which meets annually; gradually, it has broadened its concerns and has produced the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (<http://www.icmje.org>). Further details on the Uniform Requirements are reported in the Appendix.

The Uniform Requirements also report recommendations about the statistical aspects of manuscripts. The most significant recommendation is included in the initial sentence of the "Statistics" section, where the following is stated: "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the source data to verify the reported results". This statement summarizes an essential concept: the possibility of verifying the source data should warrant the reliability of both the data and the analysis performed; therefore, it should warrant the reliability of the results that were obtained.

## Misconduct Related to Data Analysis

Fabrication (invention of data or cases), falsification (willful distortion of data), not admitting that some data are missing, and ignoring outliers without declaring them constitute the main research misconduct related to statistical and data analysis [27, 28]. This is serious misconduct to the point that Buyse *et al.* [29] used the term “fraud” specifically to refer to data fabrication and falsification. Given the importance of this misconduct, its consequences on the quality of scientific information become evident. On this topic, biostatistics plays an important role and biostatisticians should be involved in preventing fraud (as well as unintentional errors), detecting it, and quantifying its impact on the outcome of the research, particularly when clinical trials are involved. In particular, the guidelines for clinical trials [30, 31, 32] indicate that a biostatistician should be involved in the protocol at all stages, from design and analysis to reviewing and in order to avoid misconduct, it is advisable that an independent biostatistician be included in the Data Monitoring Committee.

The examples mentioned so far should be considered very serious misconduct, but there are many others which can be determined by the incorrect application and/or choice of the methods used for data analysis. As an example, some of the most frequent misconduct, related to the analysis and representation of data, can be identified as follows:

- To treat missing data as zeros;
- Application of inhomogeneous statistical methods (i.e. parametric and non-parametric methods) within the same set of data;
- Application of parametric methods to data with evident non-normal distributions;
- Lack of application of specific methods in the presence of multiple comparisons;
- Erroneous interpretation of the statistical analysis results, particularly when they were obtained using sophisticated and uncommon methods of analysis;

- Omission of exact P values, i.e. reporting only the reference to the significance levels.

The most important negative effects of this misconduct can be identified in the overestimation (such as generally happens in many cases), or in the underestimation of the significant data obtained in the study, as opposed to that which would have really resulted from the application of correct techniques. Another important negative effect is the one of influencing the comparability of the results obtained in different studies; some studies might report results that are different from those reported in other studies only because the data were analyzed using incorrect methods.

In some cases, this misconduct can be generated in good faith by the scientist as a result of lack of knowledge and/or adequate statistical tools; in other cases, they might be the consequence of a deliberate choice of the researcher who prefers to report those results obtained from the analysis which provides the most significant data. In other words, in the latter example, misconduct can be the direct consequence of having chosen to apply a statistical method on the basis of its results, instead of having made the choice at the beginning (i.e.: during the protocol preparation as indicated by good statistical practice). Finally, it is important to observe that the last misconduct cited in the list reported above, might also be interpreted as the intention of masking the fact that the statistical analysis was not actually carried out.

In order to avoid this misconduct, the initial statement in the guidelines provided by the Uniform Requirements (i.e.: to allow for verification of the affirmations stated in the description of the results) becomes of fundamental importance. Unquestionably, a problem exists in relation to the practical applicability of such a recommendation; within a scientific paper, the space available for the description of statistical methods is generally not sufficient to include all the details necessary for making the analysis completely reproducible. Moreover, the

accomplishment of that which allows the direct verification of results reported in the paper would imply that a reader, or an external observer, have the possibility of gaining access to the source data; nowadays this possibility is, in reality, only theoretical for the most part and, especially for publications in print format, it is not easily applicable.

### **A Contribution of JOP to the Quality of Scientific Information**

Innovative journals, such as, for example, the one in which this article is published, can greatly contribute to the quality of scientific information, especially in relation to the problems described here. The application of the most recent innovations, in terms of new models for the creation and dissemination of scientific information, and the new possibilities provided by electronic publishing the new possibilities provided by electronic publishing technology, encourage initiatives for improving the quality of information as far as the statistical aspects are concerned.

Innovative copyright policies, where the intellectual property of scientific contributions remains with the author, and the electronic format of a publication, are fundamental to the open exchange of information in scientific communication. The first example of that is the initiative undertaken by JOP where, starting with the current issue, authors can publish their scientific papers and, at the same time, make their source data, reports of statistical analysis, as well as any other materials that they judge important, available to the scientific community in order to improve the knowledge of their results.

### **Availability of Source Data and Statistics**

The first article meeting the above criteria is the paper by Pezzilli *et al.* published in the current issue of JOP [33]; the original database and the results of the statistical analysis are freely available by means of a link in the body of the hypertext.

When a 'traditional' copyright policy is applied, such an initiative might generate new questions related to the property of the data. That is not the case for publications such as JOP where, as already mentioned, the intellectual property remains with the author who will retain the rights for conventional publication and on all additional materials about the study that he/she wishes to make available to the scientific community.

The implications of this initiative, in terms of improvement in many aspects of the quality of information and its dissemination, are very important and manifold. In relation to misconduct, the problem of the verification of the analysis performed is completely resolved, since the data and the results of the analysis are made fully available; anyone equipped with the same statistical package used by the original author can reproduce exactly the same results. In this way, much of the misconduct described in this paper can be avoided or, at least, verified and discussed by the scientific community as appropriate.

A second aspect related to the quality of the information, certainly even more important than the previous one, is directly linked to scientific knowledge: as a matter of fact, other authors have the possibility of performing further analysis on the same data. One example might be to carry out the same evaluations published in the original paper, yet applying different methodologies, or published data could be used to test new hypotheses, beyond those taken into account by the original authors. In this way, much more information can be obtained from the same data.

Other authors' evaluations of one scientist's proprietary data opens up new questions. Among them, at least two which concern the area of ethics in scientific communication, are fundamental:

- Is the permission of the original author, who owns the intellectual property of the data, mandatory for a different author in order to treat them ?
- How can we safeguard the intellectual property of the original author ?

In my opinion, the answer to the first question can only be negative. Indeed, scientific information is a patrimony belonging to the community and, as such, it must be freely accessible.

Given that data are a fundamental component of a scientific paper, they must be freely accessible and, considering their intrinsic function, to say that data must be freely accessible is the same than as saying that they must be freely evaluated.

Keeping within the limits of the dissemination of scientific results, the answer to the second question is quite evident; an author who publishes results obtained by a new elaboration on data of property of another author must clearly identify the origin and property of the data, and the citation details of the original article must be reported. The case in which commercial benefits might be obtained from the results of analysis on non-proprietary data requires discussion not appropriate in this article, since it specifically concerns the field of ethics; a 'shared' property (the original author owns the data and the scientist who performed the new analysis owns the 'idea') seems however easily applicable.

## Conclusion

In conclusion, I think that the initiative of making the source data, together with all other materials relevant to the verification of published results, available can greatly contribute to improving the quality of scientific information. It is possible to undertake such an initiative thanks to the opportunities made available in the new era of electronic publishing, the same opportunities which have allowed many other important initiatives in the last few years, particularly since an organized discussion on scientific information problems started to develop.

Finally, I hope that this initiative will be positively received by JOP authors, as well as by authors and editors of other scientific journals, so that there will be as many benefits as possible to the quality of scientific information deriving from its application.

## Appendix

The International Committee of Medical Journal Editors (ICMJE) produced multiple editions of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Over the years, issues have arisen that go beyond manuscript preparation. Some of these issues are now directly covered in the Uniform Requirements document (<http://www.icmje.org>); others are addressed in separate statements. The entire Uniform Requirements document was revised in 1997. Sections were updated in May 1999 and May 2000. The last major revision performed in October 2001 has been published in the medical literature in English [34, 35, 36, 37, 38] or in Spanish [39]. Previous revisions are also available in the literature in Portuguese [40], French [41], German [42], Japanese [43], Dutch [44], and Danish [45]. There are also on-line versions in Chinese (<http://www.wame.org/chineseuni.pdf>, 2000 edition), French ([http://collection.nlc-bnc.ca/100/201/300/cdn\\_medical\\_association/publications-f/mwc/uniform.htm](http://collection.nlc-bnc.ca/100/201/300/cdn_medical_association/publications-f/mwc/uniform.htm), 2000 edition), Portuguese ([http://www.jpmed.com.br/port/normas/normas\\_07.asp](http://www.jpmed.com.br/port/normas/normas_07.asp), 2001 edition), Russian (<http://www.mediasphera.aha.ru/trebov.htm>, 1997 edition), and Spanish (<http://www.wame.org/urmspan.htm>, 1997 edition).

---

**Keywords** Clinical Trials/statistics & numerical data; Deception; Fraud/prevention & control; Information Dissemination; Publication Bias; Reproducibility of Results; Scientific Misconduct

## Correspondence

Antonio M Morselli-Labate  
Department of Internal Medicine and  
Gastroenterology  
Alma Mater Studiorum - University of  
Bologna  
Via Massarenti, 9  
40138 Bologna  
Italy  
Phone: +39-051.549.653  
Fax: +39-051.549.653  
E-mail address: [antonio.morselli@unibo.it](mailto:antonio.morselli@unibo.it)

---

## References

1. Effect of simvastatin on coronary atheroma: the Multicentre Anti-Atheroma Study (MAAS). *Lancet* 1994; 344:633-8. [PMID 7864934]

2. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; 344:1383-9. [PMID 7968073]
3. Committee on Publication Ethics The COPE Report 1999. Guidelines on good publication practice. *Hum Reprod* 2001; 16:1783-8. [PMID 11473987]
4. Committee on Publication Ethics (COPE). Guidelines on good publication practice. *J Postgrad Med* 2000; 46:217-21. [PMID 11298477]
5. COPE: committee on publication ethics *Br J Surg* 2000; 87:1287. [PMID 11044152]
6. Committee on Publication Ethics. Committee on publication ethics (COPE): guidelines on good publication practice. *Clin Oncol (R Coll Radiol)* 2000; 12:206-12. [PMID 11005683]
7. Committee on Publication Ethics. The COPE Report 1999. Guidelines on good publication practice. *Ann Trop Paediatr* 2000; 20:87-93. [PMID 10945056]
8. COPE: committee on publication ethics. *Br J Surg* 2000; 87:837. [PMID 10931015]
9. Committee on Publication Ethics. Committee on Publication Ethics (COPE). Guidelines on good publication practice. *Dentomaxillofac Radiol* 2000; 29:195-200. [PMID 10918451]
10. Cockcroft A. COPE guidelines on good publication practice. Committee on Publication Ethics. *Occup Environ Med* 2000; 57:505. [PMID 10896955]
11. Committee on Publication Ethics Committee on Publication Ethics: the COPE report 1999. Guidelines on good publication practice. *Occup Environ Med* 2000; 57:506-9. [PMID 10896956]
12. Guidelines on good publication practice. Committee on Publications Ethics (COPE). *Br J Biomed Sci* 2000; 57:2-6. [PMID 10892026]
13. Committee on Publication Ethics: the COPE Report 1999. *Sex Transm Infect* 2000; 76:69-72. [PMID 10877609]
14. COPE: committee on publication ethics. *Br J Surg* 2000; 87:693. [PMID 10848845]
15. Doherty M, Van De Putte LB. Committee on Publication Ethics (COPE) guidelines on good publication practice. *Ann Rheum Dis* 2000; 59:403-4. [PMID 10834851]
16. COPE: committee on publication ethics. *Br J Surg* 2000; 87:265. [PMID 10718792]
17. COPE: Committee on Publication Ethics. Guidelines on good publication practice. *Br J Surg* 2000; 87:135. [PMID 10671917]
18. Committee on Publication Ethics (COPE): guidelines on good publication practice. *BJU Int* 2000; 85:2-7. [PMID 10619935]
19. COPE: committee on publication ethics. *Br J Surg* 2000; 87:6-7. [PMID 10606904]
20. Guidelines on good publication practice. Committee On Publication Ethics (COPE). *J Urol* 2000; 163:249-52. [PMID 10604369]
21. Summary of the Report of the World Association of Medical editors (WAME): an Agenda for the future *Rev Med Chil* 2001; 129:324-6. [PMID 11372302]
22. Report of the world association of medical editors: agenda for the future. *Croat Med J* 2001; 42:121-6. [PMID 11259731]
23. Ryder E. Lessons learned from the subjects treated by WAME (World Association of Medical Editors) *Invest Clin* 2000; 41:77-9. [PMID 10961043]
24. Reyes H, Kauffmann R, Andresen M. Improving the editing of medical journals and the World Association of Medical Editors (WAME) *Rev Med Chil* 1997; 125:1289-91. [PMID 9609048]
25. The creation of the World Association of Medical Editors (WAME) *Rev Med Chil* 1995; 123:536-9. [PMID 8525199]
26. Launch of the World Association of Medical Editors. *JAMA* 1995; 273:981. [PMID 7897803]
27. Smith R. What is research misconduct? COPE Report 2000.
28. Tomlinson S, Catto GRD. COPE Report 2000.
29. Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B, et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Stat Med.* 1999; 18:3435-51. [PMID 10611617]
30. Hubbard WK. ICH: Guidance on Statistical Principles for Clinical Trials (FDA Docket No. 97D-0174). *Federal Register* 1998; 63:49583-98.
31. U.S. Department of Health and Human Services. E9 Statistical Principles for Clinical Trials. FDA Guidance for Industry. September 1998.
32. U.S. Department of Health and Human Services. E6 Good Clinical Practice: Consolidated Guidance FDA Guidance for Industry. April 1996.
33. Pezzilli R, Morselli-Labate AM, Barakat B, Romboli R, Ceciliato R, Piscitelli L, Corinaldesi R. Pancreatic involvement in salmonella infection. *JOP. J Pancreas (Online)* 2003; 4:200-6.
34. Medical journal editors' uniform requirements for manuscripts submitted to biomedical journals. *J Wound Care* 2003;12:171-6. [PMID 12784598]
35. Davidoff F, Godlee F, Hoey J, Glass R, Overbeke J, Utiger R, et al. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *J Am Osteopath Assoc* 2003; 103:137-49. [PMID 12665222]

36. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. Updated October 2001. *Rom J Gastroenterol* 2003; 12:74-83. [PMID 12691084]
37. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *Ginecol Obstet Mex* 2002; 70:361-70. [PMID 12365408]
38. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals – Updated October 2001. *Mymensingh Med J* 2002; 11:61-76. [PMID 12148403]
39. Reyes H, Palma J, Andresen M. Authorship criteria for scientific publication: a translation into Spanish from the 2001 updated version of the document Uniform Requirements for Manuscripts submitted to Biomedical Journals. *Rev Med Chil* 2002; 130:1177-8. [PMID 12491837]
40. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals *Rev Lat Am Enfermagem* 2001; 9:7-16. [PMID 12040806]
41. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *CMAJ*. 1997; 156:270-85. [PMID 9012736]
42. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors *Klin Monatsbl Augenheilkd* 1993; 203(1 Suppl):1-6. [PMID 8411896]
43. Inaba Y. Introduction of 'Uniform requirements for manuscripts submitted to biomedical journals' *Nippon Koshu Eisei Zasshi* 1992; 39:297-301. [PMID 1504323]
44. The revised uniform requirements for manuscripts submitted to biomedical journals *Ned Tijdschr Geneesk* 1988; 132:2001-2. [PMID 3185804]
45. The Vancouver system. Uniform requirements for manuscripts submitted to biomedical journals *Ugeskr Laeger* 1981; 143:123-9. [PMID 722258]