

Developing and Evaluating Criteria to Help Reviewers of Biomedical Informatics Manuscripts: A Case Report

E. Ammenwerth¹, A.C. Wolff², P. Knaup², H. Ulmer³, S. Skonetzki², J.H. van Bommel⁴, A.T. McCray⁵, R. Haux⁶, C. Kulikowski⁷

¹ Research Group Assessment of Health Information Systems, University for Health Informatics and Technology Tyrol (UMIT), Innsbruck, Austria

² Dept. of Medical Informatics, University of Heidelberg, Germany

³ Institute for Biostatistics and Documentation, University of Innsbruck, Austria

⁴ Erasmus University Rotterdam, Netherlands

⁵ National Library of Medicine, Bethesda, United States

⁶ Institute for Health Information Systems, University for Health Informatics and Technology Tyrol, Innsbruck (UMIT), Austria

⁷ Dept. of Computer Science, The State University of New Jersey Rutgers, United States

Author for reprint requests and communication:

Ass.-Prof. Dr. Elske Ammenwerth

Research Group Assessment of Health Information Systems

University for Health Informatics and Technology Tyrol (UMIT)

Innrain 98, 6020 Innsbruck, Austria

phone +43 512 586734 809

fax +43 512 586734 850

email elske.ammenwerth@umit.at

<http://www.umit.at>

Abstract

Peer-reviewed publication of scientific research results represents the most important means of their communication. The authors have annually reviewed a large heterogeneous set of papers in order to produce the International Medical Informatics Association (IMIA) Yearbook of Medical Informatics. To support an objective and high-quality review process, the authors attempted to provide reviewers with a set of refined quality criteria, comprised of 80 general criteria and an additional 60 criteria for specific types of manuscripts. Authors conducted a randomized controlled trial, with 18 reviewers, to evaluate application of the refined criteria on review outcomes. While the trial found that reviewers applying the criteria graded papers more strictly (lower overall scores), and that junior reviewers appreciated the availability of the criteria, there was no overall change in the inter-rater variability in reviewing the manuscripts. The authors describe their experience as a ‘case report’, and provide a reference to the refined quality review criteria, without claiming that the criteria represent a validated instrument for quantitative quality measurement.

Keywords: publishing; criteria, selection; peer review, research; quality control; manuscripts; journalism, medical; randomized controlled trials

1 Introduction

Research is defined as carrying out an investigation into a subject or problem [1].

Communicating research results in recognized, peer-reviewed scientific journals is essential both to scientific progress, and to individual professional advancement [2]. However, as Seglen states: "Evaluating scientific quality is a notoriously difficult problem which has no standard solution." [3]. The authors' motivation to develop and evaluate quality criteria for scientific papers arose during their work in editing the Yearbook of Medical Informatics of the International Medical Informatics Association (IMIA) [4], [5]. The Yearbook appears annually and presents approximately 50 significant papers that have been published during the previous year. Its aim is to give a broad overview of the latest significant research activities in the field of health and medical informatics.

During the selection process, approximately 10,000 medical informatics papers published annually and listed in MEDLINE are reviewed and filtered to retain about 50 papers for publication in the Yearbook. Eight managing editors, each assigned to different sub-fields, first pre-select papers for review; those (approximately) 150 papers are reviewed by two external international experts, by the two editors of the Yearbook, and by the responsible managing editor. A purely quantitative review scale is used for the final review. An analysis of the external reviews of 118 papers pre-selected for the IMIA Yearbook 2001 [6] indicated a wide range of variability in "expert" scoring, with approximately one-third of the papers showing a difference in quantitative scores of 20 % or more among the two external reviewers.

The editors and managing editors attempted to refine the main review criteria with the goal of decreasing rater variability and improving quality of reviews. The authors of the current paper conducted a review of the relevant literature, developed new, refined review criteria, and conducted a small randomized controlled trial using the new criteria to assess variability and reviewers' satisfaction with the new criteria. The trial indicated that the new criteria were not useful, in an absolute sense, to create better agreement among reviewers. This is not surprising, in that reviewers with different expertise and different experiences will each view "relevant" aspects of a paper from unique perspectives.

The goal of this paper is to report the authors' attempt to define quality manuscript review criteria, and to share what was learned about their application during reviews.

2 Development of refined quality criteria for paper reviews

The authors first reviewed the available literature on review criteria for scientific articles. For example, Gunn discussed problems with the quality of electronically published clinical guidelines, such as low quality and irrelevance [7]. Elliott presented guidelines for reviewing qualitative research [9], while Jefferson assessed whether BMJ guidelines for reviewing economics submissions influenced the quality of submitted and published manuscripts in this area [10]. Also, a standard was developed for measuring quality of publications regarding randomized, controlled clinical trials (the 'Consolidated Standards of Reporting Trials', or CONSORT system) [11]. CONSORT implemented a checklist containing 21 items, covering manuscripts' methods, results, and discussion sections [11]. The German Research Association, among others, has published general guidelines for good scientific practice [12]; their recommendation 12 covers (co-) authorship, completeness of presentation of research results, and correct citation of previous work of other researchers.

Many scientific medical informatics journals provide information regarding their own quality criteria - via instructions for authors and guidelines for reviewers (e.g., BMJ, [13] and JAMIA, [14]). Review criteria of medical journals (such as BMJ) are not always directly relevant to biomedical informatics publications.

The authors' literature review showed that no previously published quality checklist met the objective of providing a comprehensive list of refined quality criteria useful for reviewing all scientific papers in medical informatics.

The authors chose a top-down approach to refining the previously used IMIA Yearbook main review criteria (with sections for: significance; quality of scientific content; originality and innovativeness; coverage of related literature; and, organization and clarity of presentation). Analysis of available literature (especially [11], [13], [14], [15], [16]) and the authors' own experiences as reviewers provided the basis for the refined criteria, which were then discussed and revised by IMIA Yearbook editors and managing editors in an iterative process.

The revised review criteria developed by this methodology are presented on the web pages of the IMIA Yearbook at <http://www.yearbook.uni-hd.de>. The revised criteria included 5 quality categories with 15 subgroups, totaling approximately 80 general questions, with approximately 60 additional questions for specific subtypes of articles. Table 1 (JAMIA online data supplement) highlights some of the differences between the review criteria of BMJ, JAMIA, and CONSORT, and compares them to the authors' refined quality criteria.

3 Evaluation of revised quality criteria and lessons learned

After developing and elaborating the revised quality criteria, the authors conducted a randomized trial, comparing the new review criteria to “standard” previous review methods (see Appendix A in JAMIA on-line data supplement for methods and results of that study). The randomized study failed to show an effect of the revised criteria on inter-rater concordance. Nevertheless, the authors found the criteria (Table 1, JAMIA online data supplement) to be helpful in the following manner.

1. The small-scale evaluation of the refined quality criteria showed that the reviewers’ absolute quality ratings fell significantly (lower scores of merit) on papers they reviewed while applying the new criteria. Reviewers commented that the refined quality criteria helped to increase their awareness of all quality criteria, so that they more easily identified weaknesses in the reviewed papers, justifying lower ratings. Stricter ratings may be a desired effect, reflecting improved review quality – especially for less experienced reviewers. More experienced reviewers may already have most of the criteria in mind.

2. An observed increased time required to apply the refined quality criteria was not surprising. In the study, all reviewers had to grade each of the (80+60) individual review criterion for every paper reviewed, so that the study could be certain that the refined criteria had been applied. During “normal” reviews, it would be expected that refined criteria would serve as useful information for novice reviewers, and as infrequently used, but available reference material for more experienced reviewers. The authors believe that the increased time required to use the refined criteria may not reduce their usefulness, as long as the criteria are available as a reference and not as an absolute requirement during review (a subject for future study).

3. Different reviewers of equal stature and ability will always judge the manuscript differently due to their varied scientific backgrounds, their variable familiarity with similar projects, and their variegated knowledge of the authors and the authors' prior work. It is therefore not surprising that the authors' randomized study found no reduction in variation between the reviewers after using the refined quality criteria.

4. The refined quality criteria do not constitute an instrument that can be used to obtain an objective quantitative assessment from reviewers, and cannot lead to a "definitive" quality score. Rather, the criteria provide a list of important considerations that can help to support thorough reviews of scientific papers (even though the results will be different due to different backgrounds of the reviewers).

5. The reviewer (or author) of a paper can use the refined criteria as a reference in case of questions, for aspects to consider when judging originality or significance of a paper. Refined quality criteria may help novice reviewers to think more thoroughly about reviews (thus perhaps leading to better reviewers). The criteria may help in assessing how, and why, to arrive at certain decisions.

6. The list of refined quality criteria is not designed to be exhaustive and will be continually revised annually by the IMIA Yearbook editorial staff.

The quality of reporting in a paper does not automatically reflect the quality of the underlying research project it describes. But a good paper makes it easier to assess the quality of a research project. Publication quality is an important aspect of research quality [15].

The authors hope that the refined review criteria will be helpful for authors of scientific papers in medical informatics, and for reviewers and editors to come to a balanced and more explicit assessment of the quality of medical informatics papers. The journal *Methods of Information in Medicine* has already adopted a draft of the refined quality criteria for its review guidelines [18].

4 Acknowledgements

The authors of this article are Editors and Managing Editors of the IMIA Yearbook 2001.

We would like to thank Martina Hutter, Steven Huesing, Thomas Kleinöder and the Schattauer Publishing Company for their support in publishing the IMIA Yearbook. We want to include the fellow Managing Editors A. Bohne, K. Ganser, C. Maier, A. Michel, V. Mludek and R. Singer in our thanks for their discussions and comments on this paper, as well as the 16 reviewers (S. Abel, B. Baumgarten, A. Bess, B. Brigl, T. Bürkle, E. Finkeissen, S. Garde, E. Lang, F. Leiner, F. Phillip, J. Pilz, U. Prokosch, M. Schwabedissen, R. Weber, T. Wendt, T. Wetter) for the participation in the study. Thanks also to Frieda Kaiser for her support, and the anonymous reviewers for their fruitful comments on an earlier version of this paper.

5 References

1. Pons. *Pons Collins Dictionary of the English Language*. Glasgow: HarperCollins. 1998.
2. de Vries J. Peer Review: The Holy Cow of Science. In: *A Century of Science Publishing* (Eds: Fredriksson E). Amsterdam, IOS Press. 2001. 231-244.
3. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997; 314(7079): 498-502.
4. van Bommel J, McCray A. *Yearbook of Medical Informatics (annual edition)*. Stuttgart: Schattauer. 1992 - 2000.
5. Haux R, Kulikowski C. *Yearbook of Medical Informatics (annual edition)*. Stuttgart: Schattauer. since 2001.
6. Ammenwerth E, Knaup P, Maier C, et al. Digital Libraries and Recent Medical Informatics Research - Findings from the IMIA Yearbook of Medical Informatics 2001. *Methods Inf Med* 2001; 40(2): 163-7.
7. Gunn IP. Evidence-based practice, research, peer review, and publication. *CRNA* 1998; 9(4): 177-82.
8. Boyer C, Selby M, Scherrer J-R, Appel R. The Health On the Net Code of Conduct for medical and health Websites. *Comput Biol Med* 1998; 28(5): 603-610.
9. Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol* 1999; 38(Pt 3): 215-29.
10. Jefferson T, Smith R, Yee Y, et al. Evaluating the BMJ guidelines for economic submissions: prospective audit of economic submissions to BMJ and The Lancet. *Jama* 1998; 280(3): 275-7.

11. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals. JAMA 1997; 277(11): 927-934.
12. DFG. *German Research Association: Proposals for safeguarding good scientific practice*. Weinheim: Wiley-VCH. 1998.
13. BMJ. Checklist for Editors and Peer Reviewers. Last accessed: Oct. 2002.
<http://bmj.com/advice/>.
14. Journal of the American Medical Informatics Association. Instructions for authors. Last accessed: Oct. 2002. <http://www.jamia.org/misc/ifora.shtml>.
15. Köbberling J. The quality of German medical journals (in German). Dtsch med Wschr 2000; 125: 1106-1108.
16. Paice E. How to write a peer review. Hosp Med 2001; 62(3): 172-5.
17. Kulikowski C, Ammenwerth E, Bohne A, et al. Medical Imaging Informatics and Medical Informatics: Opportunities and Constraints - Findings from the IMIA Yearbook of Medical Informatics 2002. Methods Inf Med 2002; 41(2): 183-9.
18. Methods of Information in Medicine. Instructions for authors. Last accessed: Oct. 2002.
<http://www.schattauer.de/zs/startz.asp?load=/zs/methods/richtl.asp>.