Reporting guidelines for health research: protocol for a cross-sectional analysis of the EQUATOR Network Library

Ferrán Catalá-López,1,2,3 Adolfo Alonso-Arroyo,4,5 Matthew J Page,6 Brian Hutton,2,7 Manuel Ridao,8 Rafael Tabarés-Seisdedos,3 Rafael Aleixandre-Benavent,5,9 David Moher,7,10

ABSTRACT

Introduction Transparency and completeness of health research is highly variable, with important deficiencies in the reporting of methods and results of studies. Reporting guidelines aim to improve transparency and quality of research reports, and are often developed by consortia of journal editors, peer reviewers, authors, consumers and other key stakeholders. The objective of this study will be to investigate the characteristics of scientific collaboration among developers and the citation metrics of reporting guidelines of health research.

Methods and analysis This is the study protocol for a cross-sectional analysis of completed reporting guidelines indexed in the Enhancing the QUAlity and Transparency Of health Research Network Library. We will search PubMed/MEDLINE and the Web of Science. Screening, selection and data abstraction will be conducted by one researcher and verified by a second researcher. Potential discrepancies will be resolved via discussion. We will include published papers of reporting guidelines written in English. Published papers will have to meet the definition of a reporting guideline related to health research (eg, a checklist, flow diagram or explicit text), with no restrictions by study design, medical specialty, disease or condition. Raw data from each included paper (including title, publication year, journal, subject category, keywords, citations, and the authors’ names, author’s affiliated institution and country) will be exported from the Web of Science. Descriptive analyses will be conducted (including the number of papers, citations, authors, countries, journals, keywords and main collaboration metrics). We will identify the most prolific authors, institutions, countries, journals and the most cited papers. Network analyses will be carried out to study the structure of collaborations.

Ethics and dissemination No ethical approval will be required. Findings from this study will be published in peer-reviewed journals. All data will be deposited in a cross-disciplinary public repository. It is anticipated the study findings could be relevant to a variety of audiences.

INTRODUCTION

Many published reports of health research studies lack clarity and transparency on how they were conducted and what they found. If published reports do not provide sufficient details about which questions were addressed, or lack descriptions of study conduct and analyses, research users may be left with an incomplete picture of what has been done; and the reliability of the findings cannot be interpreted correctly. Studies of the biomedical literature have shown that the quality of research is highly variable, with important deficiencies in the reporting of methods and results of health research studies.1–11 Unless research is adequately reported, investments are wasted because research findings are unusable.12–15 To help improve this situation, the scientific community needs close collaborations between journal editors, peer reviewers, authors, consumers and other key stakeholders.16–18

Strengths and limitations of this study

► To our knowledge, this will be the first attempt to systematically map and describe the characteristics of scientific collaboration among developers and the citation metrics of reporting guidelines of health research.

► We will not restrict inclusion based on the focus (eg, study design or medical specialty) of the reporting guideline and will be able to collect data on a broad cross-section of guidelines for reporting health research.

► Although the Enhancing the QUAlity and Transparency Of health Research Network Library is regularly updated based on comprehensive searches across the main databases indexing health-related literature, it cannot be ruled out that we will potentially miss some published reporting guidelines.

► A potential limitation could be the study will include only reporting guideline papers catalogued in one database and written in English.
Reporting guidelines of health research aim to improve transparency and quality of research reports. Reporting guidelines for health research (such as the Consolidated Standards of Reporting Trials (CONSORT) statement, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement and the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines) have been widely disseminated through publications in high-impact factor medical journals and endorsements by editors, peer reviewers and authors. In June 2008, the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network was officially launched as a global collaborative initiative seeking to improve the value and reliability of published health research by promoting transparent and accurate reporting and wider use of high-quality, robust reporting guidelines. As part of its resource development, the EQUATOR Network’s website (http://www.equator-network.org/) contains a comprehensive and up-to-date searchable database (‘the EQUATOR Network Library’) of reporting guidelines (completed and under development) relevant to health research, tools and instruments for better reporting aimed at different users and information about relevant educative activities and events.

Previous reviews have focused on the characteristics and methods for developing reporting guidelines, and the completeness of reporting health research in different areas. However, to the best of our knowledge, there has been no specific study focusing on the characterisation of research collaborations among developers of reporting guidelines for health research studies.

The objective of this study will be to investigate the characteristics of scientific collaboration and citation metrics of reporting guidelines of health research. Particularly, we will map the EQUATOR Network Library to: (1) identify reporting guidelines and key actors (authors, institutions, countries and journals) producing and disseminating them, and (2) to describe production, citations and collaboration patterns in the area.

**METHODS AND ANALYSIS**

**Design**

This is the study protocol for a cross-sectional analysis of reporting guidelines indexed in the web-based EQUATOR Network Library. The EQUATOR Network Library provides an up-to-date online collection of guidelines and policy documents related to health research reporting. These reporting guidelines are aimed mainly at authors of research papers, journal editors, peer reviewers, reporting guideline developers, educators and research funders. At present, more than 300 guidelines and related documents are available on the EQUATOR Network website (http://www.equator-network.org/). The EQUATOR Network Library provides a searchable database of all reporting guidelines (published since 1996) identified through systematic literature searches that focus on/provide specific guidance for reporting of various types of research designs, components of research reports or specific medical conditions or procedures. The EQUATOR Network Library provides guidance for different types of health research studies, for example, experimental studies including randomised trials (eg, CONSORT statement), systematic reviews and meta-analyses (eg, PRISMA statement), diagnostic and prognostic studies (eg, STARD guidelines) and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement, observational studies (eg, STROBE statement), economic evaluations (eg, Consolidated Health Economic Evaluation Reporting Standards statement), case reports (eg, COREQ guidelines, PRISMA-Protocols), animal preclinical research (eg, ARRIVE guidelines) and other reporting guidelines (eg, mixed-methods studies, reliability and agreement studies and clinical practice guidelines). In addition to these reporting guidelines, the EQUATOR Network Library lists guidance that relates to publication of health research for sections of research reports (eg, Statistical Analyses and Methods in the Published Literature for statistical methods and analysis), guidance for specific medical conditions, procedures or clinical areas (eg, cardiovascular medicine, clinical oncology, paediatrics, public health or genetics), but also reporting guidelines under development. Currently, the EQUATOR Network Library website lists only references to journals that published the reporting guideline first, not later republications (often in a national or local journal, usually in a local language edition).

**Eligibility criteria**

To be eligible for inclusion, published papers will have to meet the following definition of a reporting guideline related to health research: ‘A reporting guideline is a checklist, flow diagram or explicit text to guide authors in reporting a specific type of research, developed using explicit methodology’. We will include reporting guidelines with no restrictions by study design, medical specialty, disease or condition. We will include published papers of reporting guidelines (ie, ‘completed’ reporting guidelines) written in English. Translations of any existing reporting guideline in other non-English languages will be excluded (eg, CONSORT translation into Spanish or Chinese). Meeting abstracts, proceedings paper (journals, book based), editorials, book chapters, corrections, retracted publications and other items (eg, data paper, news and notes) will be excluded. Reporting guidelines under development will be also excluded.

---


Open access
To provide a reliable summary of the reporting guideline literature, we will search for completed reporting guidelines indexed in the EQUATOR Network Library from inception to 31 July 2018 (this date will represent 10 years after the official launch of the EQUATOR Network in 2008). We will first identify all the published papers catalogued in the EQUATOR Network Library. For all reporting guidelines, an information scientist will obtain from the EQUATOR website the following: the title (or reference), standard identifier for each paper (eg, PubMed ID or a digital object identifier, if the paper has one), study design and date of record last updated. Additionally, we will contact the EQUATOR Network Library team for clarifications and further data (if necessary). To ensure consistency in the searching process, a second researcher will perform complementary hand-searches (eg, websites of reporting guidelines catalogued in the EQUATOR Network Library) (see examples in table 1) to check for potential indexing omissions or errors. We will use all the standard identifiers retrieved for published papers to carry out a search in PubMed/MEDLINE in order to have access to full references, titles and abstracts. With this information, we will perform a search of the Web of Science (Clarivate Analytics, Philadelphia, Pennsylvania, USA) based on the title or the standard identifiers of the papers. Intended date for the searches in PubMed/MEDLINE and Web of Science will be 15 April 2019. The search results will be imported into a Microsoft Access (Microsoft, Seattle, Washington, USA) database.

Screening and selection
All titles and abstracts will be screened against our eligibility criteria by one researcher and verified by a second researcher. We will retrieve the full text of any paper(s) meeting our eligibility criteria or for which eligibility remains unclear. Any discrepancies in the screening of titles and abstracts and selection of full-text papers will be resolved via discussions by the research team, if necessary.

Data extraction and data normalisation
Raw data from each paper will be exported (downloaded online) from the Web of Science (Clarivate Analytics) by one researcher (senior information specialist). Data will be collected in Microsoft Access (Microsoft) containing the following: title, publication year, journal, subject category, keywords, citations (the number of times a paper has been cited by other publications), source, language, document type (eg, article or review according to Document Type Field from the Web of Science) and the authors’ names, author’s affiliated institution (eg, institution name, department and institution address including city), and country. All data will be verified by a second researcher to minimise potential omissions or information errors. Any discrepancies in the extracted data will be resolved via discussion or adjudication by a third researcher if necessary.

To ensure consistency in the data extraction, a process of data cleaning and normalisation will be conducted by two members of the research team to bring together the different names of an author, institution, country, journal and keywords. Specifically, one researcher will check the names by which an individual author appeared in two or more different forms (eg, ‘Douglas G Altman’ or ‘Douglas Altman’ or ‘Doug Altman’), using coincidence in that author’s affiliated institution as the basic criterion for normalisation (eg, Oxford University, UK). In the case of author’s affiliated institution, names in many records may include two or more institutions (eg, university hospitals, research centres and academic institutions). So, we will proceed to distinguish between institution names by recording all variations of any individual macroinstitution as could be identified for each record (eg, for the institutional address ‘The Ottawa Hospital and University of Ottawa, Canada’, the normalisation approach will be to present ‘University of Ottawa, Canada’ separately from ‘The Ottawa Hospital, Canada’). Missing data (eg, missing country information of an affiliated institution) will be substituted and completed by manual web search.

Keywords (‘author keywords’ and ‘keyword plus’) which are automatically assigned by the Web of Science (Clarivate Analytics) from the title of the paper will be extracted and used to represent the conceptual content of the publications.

Typographical, transcription and/or indexing errors will be corrected or removed. All potential discrepancies will be resolved via discussion.

Data analysis
The analysis will be descriptive, with data summarised as frequencies and percentages for categorical variables or median and IQR for continuous variables.

Our primary analysis of production, citations and collaboration patterns will be based on single papers (paper-level analysis). The description of the scientific production and citation impact will be carried for authors, institutions, countries and journals (eg, total number of papers and total number of citations). Tables will be used to present the most prolific authors (5 or more published papers), institutions (5 or more papers), countries (10 or more papers), journals (10 or more papers) and the most cited papers and reporting guidelines (‘top 50 citation classics’).

We will describe collaboration patterns with respect to authors, institutions and countries: number of signatories (total number of authors included in all the papers of each author), number of collaborators (and main collaborators in terms of number of collaborations), number of papers with coauthors, collaborations between authors (‘interindividual collaboration’), and collaboration index which is the mean number of signatories per paper. We will identify collaborations between institutions (‘institutional collaboration’) and countries (‘international collaboration’). For institutions, we will differentiate between institutional collaborations in the same country.
Network analyses will be carried to study the structure of collaborations between authors, institutions and countries. Collaboration networks will be constructed by identifying the regularities in the patterning of relationships among entities (eg, which might be authors, institutions and countries) for each single paper. We will present network diagrams to illustrate the structure of scientific collaborations. A network diagram is a visual representation of a collection of nodes or vertices (eg, numbers of papers or collaborations) and the relational ties connecting them, called links or edges (eg, numbers of papers or collaborations).
Particularly, we will identify clusters of more intense collaboration networks by applying a priori thresholds (‘collaboration intensity’) based on the numbers of papers signed in collaboration between authors (5 or more papers), institutions (5 or more papers) and countries (10 or more papers). All the network diagrams will be presented as undirected-weighted graphs (where node sizes are proportional to the number of papers, links are bidirectional, and the link thicknesses are proportional to the number of collaborations). Key standard measures and metrics for quantifying network structure (centrality measures such as ‘degree of centrality’, ‘closeness’ and ‘betweenness’) will also be calculated to complement the cluster analysis. Centrality measures identify the most prominent (‘key’) members of a scientific community, that is, those who are extensively involved in relationships with other network members. Degree of centrality is a simple centrality measure that counts how many connected links a node has (eg, the number of coauthors of a given author). If a node has many ties compared with nodes, this indicates that this node has a central position in the network. In a collaboration network, it seems reasonable to suppose that a node (eg, author) having connections to many others might have more activity or more prestige than those having fewer connections. Centrality can also characterise the shape of a whole network. Closeness is based on the notion of distance. Closeness measures the mean distance from a node to other nodes (eg, an author that has high centrality closeness is likely to receive information more quickly than others in the network). Closeness measures efficiency or independence. Betweenness measures the extent to which a node (eg, author) lies on paths between other nodes or how much a node connects other nodes with each other (eg, an author who is high in betweenness is able to control the flow of information or resources passing between other coauthors).

We will conduct topical (also called linguistic, semantic or textual) data analyses and visualisations to determine the topic coverage of a body of text. The most frequently used keywords will be identified and presented for each journal. Word clouds displaying the frequency of terms (with larger words depicting the frequency of occurrence) will be generated for the main keywords to identify major topics. We will also present the ‘co-words network’ of keywords representing the co-occurrence phenomenon of highly frequent words in the papers.

Multiple publication of identical paper(s)
It is important to note that for a paper presenting a reporting guideline may have been published simultaneously by multiple journals to support wider dissemination and uptake. Some of these reporting guidelines published in multiple journals include CONSORT statement and PRISMA statement. Secondary analysis will consider and characterise production, citations and collaboration patterns by integrating data for multiple (identical) papers of a single reporting guideline (guideline-level analysis). This will involve the following: (1) authorship and collaboration data will be recorded once and (2) citation impact will be computed by grouping the total number of citations across the multiple (identical) papers published in different journals.

Additional analyses
We will undertake a subgroup analysis to compare papers and reporting guidelines in terms of scientific productivity, citation and collaboration patterns before and after the official launch of the EQUATOR Network in 2008: ‘before 2008 (pre-EQUATOR Network)’ versus ‘after 2008 (post-EQUATOR Network)’.

Depending on the productivity and collaboration patterns, we plan to explore whether sensitivity analyses related to thresholds (‘collaboration intensity’) are feasible using different numbers of papers in collaboration (eg, between authors, institutions and countries) to see how stable our results are.

Software considerations
Relational structure (network) visualisations and metrics will be produced by using PAJEK (University of Ljubljana, Slovenia), a free software package for large network analyses. Word clouds will be drawn using Wordle, a free online tool, for graphical representation of words frequency. We also plan to explore additional visualisation tools, considering this is a constantly evolving field.

Patient and public involvement
No patients and/or public were involved in setting the research question, nor were they involved in developing plans for design (or implementation) of this study protocol. No patients and/or public will be asked to advise on the interpretation or writing up of results. There are no specific plans to disseminate the results of the research to the patient community.

Ethics and dissemination
To the best of our knowledge, this cross-sectional analysis will be the first attempt to systematically map the content of the EQUATOR Network Library using methods and tools from network science. The study will serve to identify reporting guidelines of health research, key actors producing and disseminating them, citation impact and scientific collaborations in the area. It will be helpful to determine the most intense collaboration patterns, most productive authors, institutions and countries developing reporting guidelines, most prolific journals disseminating them and the ‘citation classics’ of guidelines for reporting health research studies.

Any amendments made to this protocol when conducting the analyses will be outlined and reported in the final manuscript. Findings from this study will be published in peer-reviewed journals. All data underlying the findings reported in the final manuscript will be deposited in a cross-disciplinary public repository, such as the Open Science Framework or Zenodo.
We anticipate the study could be relevant to a variety of audiences including journal editors, peer reviewers, research authors, guideline developers, research funders, educators and other potential key stakeholders. Moreover, the study findings could further be used to strengthen scientific capacity for international, national and subnational collaborations in order to increase value and reduce waste from incomplete or unusable reports of health research studies.  

**Author affiliations**

1. Department of Health Planning and Economics, National School of Public Health, Institute of Health Carlos III, Madrid, Spain
2. Knowledge Synthesis Group, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
3. Department of Medicine, University of Valencia/INCLIVA Health Research Institute and CIBERSAM, Valencia, Spain
4. Department of History of Science and Documentation, University of Valencia, Spain
5. Unidad de Información e Investigación Social y Sanitaria-UISYS, University of Valencia, Spain
6. Instituto Aragonés de Ciencias de la Salud (IACS), Red de Investigación en Salud, Zaragoza, Spain
7. School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
8. School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia
9. School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
10. Instituto Aragonés de Ciencias de la Salud (IACS). Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Zaragoza, Spain
11. Ingenio-Spanish National Research Council (CSIC) and Universitat Politècnica de Valencia (UPV), Valencia, Spain