

*This group comprises eight authors

Indrayan A,
Vishwakarma G¹,
Malhotra RK², Gupta
P³, Sachdev HPS⁴,
Karande S⁵, Asthana
S⁶, Labani S⁶

Clinical Research,
Max Healthcare,
New Delhi; ¹ Biostatistics:
Indian Spinal Injuries
Centre, New Delhi;
² Surgical Oncology, Dr.
B. R. Ambedkar Institute
Rotary Cancer Hospital,
All India Institute of
Medical Sciences,
New Delhi; ³ Pediatrics,
University College
of Medical Sciences,
Delhi; ⁴ Pediatrics and
Clinical Epidemiology,
Sitaram Bharta Institute
of Sciences and
Research, New Delhi;
⁵ Pediatrics, Seth G.
S. Medical College
and KEM Hospital,
Mumbai, Maharashtra;
⁶ Epidemiology
and Biostatistics,
National Institute of
Cancer Prevention and
Research (Indian Council
of Medical Research),
Noida, Uttar Pradesh,
India

Address for correspondence:

Dr. Indrayan A,
E-mail: a.indrayan@gmail.
com

Received : 04-06-2022
Review completed : 12-09-2022
Accepted : 13-09-2022
Published : 04-11-2022

The development of QERM scoring system for comprehensive assessment of the Quality of Empirical Research in Medicine - Part 1

Research Quality Improvement Group*

ABSTRACT

Purpose: Whereas a large number of features are mentioned to connote the quality of medical research, no tool is available to comprehensively measure it objectively across different types of studies. Also, all the available tools are for reporting, and none includes quality of the inputs and the process of research. The present paper is aimed to initiate a discussion on the need to develop such a comprehensive scoring system (in the first place), to show that it is feasible, and to describe the process of developing a credible system.

Method: An expert group comprising researchers, reviewers, and editors of medical journals extensively reviewed the literature on the quality of medical research and held detailed discussions to parse quality at all stages of medical research into specific domains and items that can be assigned scores on the pattern of quality-of-life score.

Results: Besides identifying the domains of the quality of medical research, a comprehensive tool for scoring emerged that can be possibly used to objectively measure the quality of empirical research comprising surveys, trials, and observational studies. Thus, this can be used as a tool to assess Quality of Empirical Research in Medicine (QERM). The expert group confirmed its face and content validity. The tool can be used by the researchers for self-assessment and improvement before submission of a paper for publication, and the reviewers and editors can use this for assessing the submissions. Published papers can also be rated such as those included in a meta-analysis.

Conclusion: It is feasible to devise a comprehensive scoring system comprising domains and items for assessing the quality of medical research end-to-end from choosing a problem to publication. The proposed scoring system needs to be reviewed by the researchers and needs to be validated.

KEY WORDS: Empirical research, medical research quality, QERM score, quality assessment, scoring system, tool to assess quality

Access this article online	
Quick Response Code:	Website: www.jpgmonline.com
	DOI: 10.4103/jpgm.jpgm_460_22
	PubMed Id: 36348606

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Research Quality Improvement Group. The development of QERM scoring system for comprehensive assessment of the quality of empirical research in medicine - Part 1. J Postgrad Med 2022;68:221-30.

Introduction

The concept of quality generally refers to the inherent properties of a process or product that meet the stated objectives. In the case of medical research, the objective generally is to find new ways to improve health. This is mostly achieved by empirical research. This kind of research is an inherently imperfect endeavor in any field, but medical research is much more inflicted with uncertainties because of its interface with volatile humans. Thus, it requires special tools for assessing quality.

The quality of medical research has been in discussion for quite some time^[1-4] but the concern has steeply increased in the past decade as the number of papers and journals has exponentially increased. Many of these publications are believed to be of dubious quality.^[3] While discussing “scandal” in medical research, Altman^[5] stated that poor quality is widely acknowledged, but it raises minimal concern among the medical professionals. Ioannidis^[6] commented on false findings in much published research and drew attention to the flawed and misleading findings in many publications resulting in huge wastage of resources.^[7] Chalmers and Glasziou^[8] also highlighted this enormous wastage. Thus, the quality of medical research needs urgent attention—not just for publication, but also the process, so that quality research is conducted and reported.

The quality of medical research has been interpreted differently by different workers. Ioannidis^[1] emphasized truthfulness of results and Mendoza and Garcia^[2] expressed concern with reproducibility. ESHRE Capri Workshop Group^[3] showed concern with credibility and utility, and Mische *et al.*^[4] discussed scientific rigor and transparency. With such diversities, it is important that a consensus is evolved. The core domains of quality need to be identified and the components of each domain are specified for assessing specific ingredients of quality.

Most comments in the literature on the quality of medical research are based on the publications rather than the actual process. However, the quality assurance should be done all through the process, and not just at the end, to ensure the excellence in the inputs, the process, and the final product.

An expert Research Quality Improvement Group, comprising researchers, reviewers, and editors of medical journals, held intensive consultation among themselves and considered a wide spectrum of features that constitute “quality.” The endeavor was to include not just the quality of the written draft but also the inputs and the process. The Group explored the possibility of developing a scoring system comprising domains and items that can measure the quality of medical research in a comprehensive and objective manner, much like the quality-of-life index. The deliberations were limited to the data-based empirical research because of the widely divergent issues in other kinds of research. They resulted in identifying the domains of the quality of medical research. Based on these domains, a proposal emerged for a tool to assess Quality of an Empirical Research in Medicine (QERM) that can measure the quality of the research

process and the output. Evidence from CONSORT, STROBE, and STARD indicates that such guidelines do help in improving the quality of reporting.^[9-12] Adherence to the items of our proposed scoring system may help in improving not just the reporting but the entire process of medical research.

The Group realized at the outset that it is not feasible to observe the research process actually followed by different workers in their institutions and organizations, and the only effective way to assess the quality of medical research by a third party is by evaluating the written document that describes the research. We included the quality of steps at the planning and execution that can be used for self-assessment and did not restrict to the quality of manuscripts before publication when peer review is routinely done. The proposed scoring system contains items that could be useful to frame a worthwhile protocol, and to adopt an appropriate process of conducting quality research, including the steps beginning with the choice of the topic and ending with the publication of a paper. Reporting also is a part of this unique scoring system. The scoring may be useful for the reviewers and editors also to assess the quality of a submission that generally includes the statements on the conception and the process. Funding agencies can also use this scoring tool to assess the quality of the proposals and the reports of the concluded project. Published papers can also be assessed, including those in a meta-analysis.

This communication reports the process adopted and the progress made to quantify the quality of empirical medical research through scoring. This includes a checklist. We follow the dictum that excellence should not be the enemy of good.^[13] The objective is to initiate a discussion on the need to develop such a system (in the first place), to show that it is feasible, and to propose a credible scoring system to assess the quality of medical research for review by the researchers.

Materials and Methods

To comprehensively capture the traits that reflect the quality of medical research, the following activities were undertaken by the Group.

- (i) Extensive review of the literature concerned with the quality of medical research to identify potential traits that could go into developing a scoring system. For this, PubMed database was searched for articles containing the terms (“research quality” OR “quality of research”) in the title published during last ten years and has full text available. This yielded 84 articles (as of January 22, 2022). This could be a restricted sample but may have provided a reasonable snapshot of the recent thinking on this issue.
- (ii) Since most journals around the world depend on the inputs of the reviewers to decide to publish or decline, reviewers’ guidelines of several journals were consulted including PloS Reviewers Guidelines^[14] and Wiley Guidelines.^[15]
- (iii) The origin of CONSORT,^[16] STROBE,^[17] and STARD^[18] guidelines was studied to learn lessons regarding how to begin and complete such an exercise, and to pick features for assessing the quality of a research.
- (iv) The SQUIRE guidelines^[19] and ICMJE recommendations^[20]

- for improved reporting were examined.
- (v) Other quality assessment tools such as OQAQ^[21] for review articles, AMSTAR^[22] for systematic reviews, QUDAS^[23] for diagnostic tools studies, NOS^[24] for observational studies, and COREQ^[25] and the one proposed by Mays and Pope^[26] on qualitative research were studied.
 - (vi) Other relevant articles we could locate on this topic through internet search and cross-references were also studied. The attempt was to comprehensively capture all possible aspects of quality applicable to all levels of empirical research from observational studies to clinical trials.

The articles consulted by us are in the reference list and the Supplementary Material - 1. This also includes a separate list of articles on describing the process of developing tools such as checklists and scoring systems without validation at that stage.

Because of enormous and wide variety of information, the Group decided to list all the terms that connote the quality and bin them into suitable domains, such that the terms within each domain are related, but are largely unrelated with the terms in the other domains. The domains are like abstract constructs, whereas the items are the observable entities.

Domains of quality of medical research

The above-mentioned review identified more than 50 terms that could connote different aspects of the quality of medical research. These were classified using established norms.^[27] In the context of empirical medical research, these may be understood as adequacy of the process, and truthfulness and applicability of the results and the conclusions, which comes from clarity regarding the research question and the methodology of the investigation.^[28,29] This was extended to choosing the problem, the thought process, and the execution of research, besides the draft of the manuscript. The five domains thus identified are in column 1 of Table 1.

Table 1: Domains of the quality of medical research and their constituents

Domain	Constituent items
1	2
Clear	Fully identified, open, transparent, unambiguous, well defined
Adequate	Comprehensive, complete, enough, ethical, focused, full, justified, measurable, novel, original, plausible, rational, reliable, repeatable, replicable, reproducible, rigorous, significant, sufficient, worthy
Truthful	Accurate, adequate level of evidence, appropriate, believable, beyond doubt, credible, correct, convincing, factual, integral, objective, real, reasonable, right, trustful, unbiased, valid
Applicable	Accessible, affordable, beneficial, convenient, cost-effective, feasible, generalizable, harmless, impactful, important, interpretable, pragmatic, relevant, robust, scopeful (with wide scope), setting (community, hospital, clinic), simple, sustainable, timely, translationable, understandable, useful, valuable
Reporting	Statement of all the above: Articulate, brief, coherent, complete, concise, focused, follow guidelines, logical, organized, succinct, understandable

Brainstorming was done to bin these 50-odd terms into the identified domains. One person (AI) was asked to come up with a classification for review by the Group. After some discussion and modification, the agreed classification is as shown in column 2 of Table 1. This classification helped to define each domain so that it has the same meaning for everybody. It also helped to understand what each domain contains and what does it represent.

The first four domains comprise the traits to be considered mostly at the time of planning and execution, whereas the fifth is on drafting the manuscript. The draft of the paper should contain the statements on the first four domains plus more as mentioned later. The following are the briefs of each domain. The details are provided in the Supplementary Material - 2.

The items numbered under each domain in the following paragraphs are based on the aforementioned review of the literature and the Group discussion regarding various components of research. The components begin with the choice of the research question and end with the drafting of the report.

Clarity

Scott-Findlay and Pollock^[30] called for conceptual clarity and Shaw^[31] discussed clarity in research integrity. An important ingredient of clarity is transparency.^[4,32-35] A brief of clarity, including transparency, regarding various components of medical research, is as follows: The details are in the Supplementary Material - 2. (i) Unambiguous problem, (ii) complete specification of the objectives, (iii) full clarity regarding the target population, (iv) clarity regarding the design of the study, (v) complete specification of the sample size, considering the exclusions and dropouts, (vi) full specification of the intervention, if any (vii) clearly formulated tools for data collection, (viii) clear process of elicitation of information, (ix) road map for analysis of data, (x) visualization of the expected results, and (xi) fitting negative and positive results in the research jigsaw puzzle.

Adequacy

Adequacy is sufficiency without overdoing. Bordage^[36] and Pierson^[37] considered inadequacy of contents as a major cause of rejection of manuscripts.

The “adequacy” domain contains many traits—most talked about is reproducibility.^[2,38-41] This is generally understood as the ability of a competent researcher to get nearly the same result using similar material from another setup when similar methods are used as by the original investigator.^[42] This is different from replicability^[43] and repeatability but can also be broadly considered parts of reproducibility. According to Goodman *et al.*,^[44] reproducibility incorporates features of design, reporting, analysis, interpretation, and corroborating studies.

The adequacy of different components of research can be enumerated as follows: (i) original, novel, and justified research question, (ii) sufficient resources for completing the research, (iii) measurable objectives, (iv) adequate intervention

to achieve the stated objectives, (v) appropriate target population, (vi) adequate tools for collection of the data, (vii) no relevant variable missed, (viii) study design that takes care of confounders and interactions, (ix) sufficient sample size not discounting the advantages of small samples,^[45] (x) representative sample, (xi) ethical and complete data collection, (xiii) appropriate analysis for the type of data and focused on the research question, (xiii) the results with sufficient reliability, and (xiv) valid reasons available if the research question is not fully answered. For details, see the Supplementary Material - 2.

Truthfulness

Reproducible research can have bias when the same bias recurs every time the research is reproduced. A comprehensive list of biases is given by Indrayan and Holt.^[46] Validity is the ability to reach the truth with no contrarian example. Since empirical research can never include future subjects, the veracity of the hypothesis cannot be determined, and truthfulness is assumed when its falsehood cannot be demonstrated.^[47] Ioannidis^[1,6] emphasized “truthfulness” of results as the core component of the quality of medical research.

The following brief incorporates the “truthfulness” in the context of various components of research. The details are in the Supplementary Material - 2. (i) Accurate research question anchored with prior evidence, (ii) objectives directly related to the research question, (iii) target population is specified, (iv) variables chosen provide a correct answer to the research question, (v) valid intervention, (vi) chosen end-points provide answer to the research question, (vii) valid tools for eliciting the data, (viii) appropriate subjects of the study, (ix) accurate measurements, (x) design of the study provides unbiased results, (xi) reliability and power of the study are medically relevant, (xii) correct method of analysis with no *P*-hacking,^[48] (xiii) credible and evidence based results, (xiv) internally and externally validated results, (xv) chance of false results minimized, (xvi) factors other than data that can affect the results are considered,^[49] (xvii) conclusion considered corroborative evidence, and (xviii) imperfections in the tools and alternative explanations considered for the conclusion.

Applicability

There may be isolated examples of conclusions that are applicable but not useful, or vice versa, but these two traits generally go together. Limitations, both known and unknown, can hamper the applicability. Any medical research is considered good if its results and conclusions are useful for improving the health of people directly or indirectly. A clear, truthful, and reproducible research, described in the preceding sections, does not necessarily imply that it is useful too. Utility relates to the extent to which the results are going to impact the practice^[50] and stands on its own as a domain of quality. The importance of the applicability has been emphasized by Goodman *et al.*^[44] and Ioannidis.^[51]

Research component-wise applicability and utility comprise (i) useful and relevant questions, (ii) the research is timely and

objectives amenable to translate into practice, (iii) study setting specified, (iv) inclusion and exclusions criteria not too restrictive, (v) how target population would benefit from the research, (vi) effect size is medically significant, (vii) control group on existing regimen rather than placebo, (viii) the intervention easy to adapt to local conditions, (ix) chosen variables work under varying conditions, (x) implementation of results feasible under varying conditions, (xi) the methodology can be adopted by other workers to check the replicability, (xii) robust results, and (xiii) the conclusions have demonstrable applicability. The details of each of these are in the Supplementary Material - 2.

Reporting

Research generally culminates into a report for dissemination of the findings to the target audience. A huge project may require a full volume, but many reports are published in a journal in a summarized but self-contained paper reflecting the entire process of research and the methodology to reach a conclusion. This section is restricted to the quality of the draft of a paper for publication in a journal.

Many articles have appeared that advise on how to write a paper for publication.^[52-54] Most of these have implications for the planning and execution also. The guidelines such as CONSORT,^[10] STROBE,^[11] and STARD^[12] are primarily for the content and style of the drafting of specific types of studies, and SAMPL^[55] is for statistical reporting.

The basic tenet of quality reporting is that the text describes the full process of research, including the results, and stated in a focused, concise, and precise manner without losing clarity. The draft should be well organized and follow a logical format such as IMRaD.^[20]

Many of the following suggestions for quality of a draft may look like a repetition of what we advised earlier in this communication, but the earlier advice was for conception, thought process, and the execution, whereas the advice now is on stating all of that in the draft of the paper. The possibility of an excellent thought process but poor reporting cannot be ruled out. In a way, the following reinforces the advice provided earlier for various domains: (i) accurate title that describes the research, (ii) unambiguous research question, (iii) complete specification of all the variables, (iv) clear identification of the target population, (v) factual design and the guidelines concerned with the design followed, (vi) target and actual sample size with justification, (vii) methods used to elicit the data, (viii) what data collected from whom, and adequacy and correctness of the available data, (ix) complete method of the analysis, (x) SAMPL guidelines^[55] followed for statistical reporting, (xi) complete results including the unfavorable ones, (xii) limitations kept in view while interpreting the results, (xiii) the conclusion fully answers the research question, (xiv) all relevant references cited, (xv) data sharing considered for others to replicate, (xvi) supplementary material for fuller explanation, and (xvii) keywords adequately describe the thrust of the research. The details of each of these are in the Supplementary Material - 2.

Results

For a scoring system to be applicable, it must not be too long, and practically feasible and manageable. Thus, only the essential items are included in our proposed scoring without leaving out any aspect that substantially affects the quality [Table 2]. This scoring gives more weight to the process (methodology) although the outcome (result) also gets its due share. Each item is given a score 2 for nearly full satisfaction, score 1 for partial satisfaction, and score 0 for almost no satisfaction. The scores for each domain can be assigned standalone to quantify the quality of different aspects of empirical research on the pattern of QoL, or can be used for self-assessment by the researcher. Higher weight to more important items was considered but was disfavored as it could introduce complexity and subjectivity. Instead, the number of items in various domains was different and this number matched fairly well with our opinion regarding the importance of that domain. For example, the highest number of items (13) is for “Adequacy” and “Truthfulness” followed by “Clarity” (10 items). The total score is the simple sum, considered valid for reflective scoring models^[7] opposed to formative measurement models with a combination of heterogeneous items.^[56] In case categories are preferred, we suggest considering a total score of less than 50 as poor, 50–74 as tolerable, 75–89 as good, and 90+ as excellent. The utility of this scoring system is also in assessing which quality domain of the research is strong and which domain is weak. The items in column 3 of Table 2 can also be used as a checklist without scoring.

The extract in Table 3 is the QERM-Brief for the reviewers who have access to only the manuscript and are generally hard pressed for time. In this case, the items are to be scored as 1 for “satisfied” and 0 for “not satisfied.” Presentation in terms of “concise, coherent, unambiguous, clear, and understandable, and supported by tables and graphs, and follow the reporting guidelines” is given a maximum score of 10 and the statements regarding all other domains together have a maximum score of 40. The editors can devise their own grouping but, in our opinion, the manuscripts with a total score of less than 30 can be rejected straightway, 30–39 as requiring revision, and 40+ as acceptable, provided there is no specific concern of the reviewer. The scoring may be able to provide a holistic assessment, largely free of the specialized field of the reviewer. This can minimize the scope for unbalanced review. Unethical research can be discarded straightway.

The experts in the Group verified face and content validity of the scoring systems after intense discussion. The other kinds of validity (including external validity and construct validity) are under process and will be reported separately. This communication is limited to the process we followed and the progress we made toward quantifying the quality of empirical medical research.

Discussion

Scoring in clinical practice has brought in relief and ease of evaluation in much the same way as the measurement of laboratory parameters has brought in medical care. Although the perception of the evaluator cannot be completely ruled

out, a scoring system can achieve objectivity to a large extent because all the items are fully specified. In the past, scores such as for quality of life, APACHE for critical care patients, and APGAR for newborns have found wide applications. We expect that the domains and items we list for QERM scoring will help researchers not just in assessing their research but also in making them aware of the features that need to be considered while planning and executing a research project.

Besides the primary purpose of helping to plan, execute, and report quality research, the QERM scoring sheet can serve many other purposes. As discussed by Sandercock and Whitley^[57] for quality research, the QERM can help the emerging generation of researchers to get started on a sound footing, create teaching resource, and help clinicians working in resource-poor settings to optimally allocate resources for quality research. Publications and funding proposals can also be evaluated, including the articles in a systematic review/meta-analysis.

Jadad *et al.*^[58] listed 49 items such as random allocation and blinding that connote quality and developed a scale for quality of reporting of RCT with a focus on control of bias. Catillon^[59] analyzed more than 20,000 RCTs and used criteria such as “adequate methods” and “poor reporting” for assessing quality. Higgins^[60] discussed the Cochrane Collaboration tool for assessing bias in RCTs. Olivo *et al.*^[61] reviewed nine scales used in physical therapy trials and 16 scales used for other areas of healthcare research. All these are focused on specific areas such as health services, dental injuries, or acupuncture. Many of these scales follow the pattern of items or domains and sub-items.

Gabriel and Maxwell^[62] emphasized on the “level of evidence,” which is an indicator of the potential for bias.^[63] Montagna *et al.*^[64] were of the view that “adoption of standardized protocols” can foster the strengthening of scientific publications. For Ueda *et al.*,^[65] the impact factor of the journal is the main consideration for assessing quality! Glasziou *et al.*^[66] mentioned reliability of answers to important clinical questions as an important component of the quality of medical research. We believe our QERM score is comprehensive and contains all these traits of quality.

Most guidelines require that the checklist be submitted along with the paper for publication without scoring system. But some use scoring system. R-AMSTAR^[22] quantifies quality for systematic reviews, QUADAS^[23] for diagnostic tools studies, NOS^[24] for observational studies, and Jadad scale^[58] for RCTs. Our effort is to devise a common scoring system covering several kinds of studies. Second, the existing tools mostly focus on reporting and not much on the conceptual framework and the process. The proposed QERM is more comprehensive with wider coverage, including aspects of conceptualization and the process of research.

We realize that the proposed QERM scoring is complex but that is so because it is comprehensive and incorporates several facets of medical research. The researcher may have to first understand each term signifying the quality [Table 1] of different aspects of the research before using the QERM score.

Table 2: Scoring system for assessing the quality of empirical medical research: from the conceptualisation to the publication - Detailed QERM scoring system (for the researcher)

Explanation of the domain	Domain	Item	Score or check
1	2	3	4
This section is only for CLARITY and not for adequacy, truthfulness, and applicability. This answers WHAT of the study and includes transparency, openness, unambiguity, well defined, and fully identified.	1 CLARITY How much are you satisfied with the following? (Scores: Full=2, Partial=1, No=0)	1.1 Research question and the objectives have no ambiguity 1.2 Target population is well defined 1.3 Variables under study (antecedents, intervention, confounders, and outcome) and their definition are clear 1.4 Design of the study (prospective/retrospective/cross-sectional, allocation, blinding, etc.) and flow of the cases is clearly conceptualized and is transparent 1.5 Method of selection of the study subjects is clear, eligibility clearly specified 1.6 Tools and data elicitation methods (observation, interview, investigation, examination-even for records-based study) are clear 1.7 Data obtained and available for analysis are clear 1.8 Methods of analysis and the road map are clear 1.9 Results expected or obtained are clear 1.10 Conclusions obtained or visualized are clear considering the negative and positive results	
	Total score for CLARITY (maximum 20)		
This section is only for ADEQUACY and not for clarity, truthfulness, and applicability. Adequate includes comprehensive, complete, enough, ethical, focused, full, justified, measurable, novel, original, plausible, rational, reliable, repeatable, replicable, reproducible, rigorous, significant, sufficient, worthy	2 ADEQUACY How much are you satisfied with the following? (Scores: Full=2, Partial=1, No=0)	2.1 Research question is adequate (original, requires investigation) 2.2 Adequate resources (facilities, expertise) 2.3 Objectives are measurable 2.4 Chosen study variables are adequate to answer the research question and are measurable 2.5 Ethical standards are met 2.6 The reliability/power is stated for calculating the sample size for each group 2.7 The study design is adequate (capable of obtaining all the relevant data on antecedents, intervention, confounders, and outcomes on a representative sample), including follow up, if any 2.8 Sufficient details of the methodology are available for replication by others 2.9 Sufficient data are available after exclusions of missing values, outliers, etc., and exclusions properly accounted 2.10 All the required methods of analysis have been used (no relevant method ignored)- different types of analysis done - the analysis is rigorous with assumptions duly verified 2.11 Reliability of results is adequate (sufficient precision to generate confidence) and the presence of medically significant effect established or denied with reasons 2.12 The plausibility of the results is demonstrated 2.13 The answer to the research question is full and convincing	
	Total score for ADEQUACY (maximum 26)		
This section is only for TRUTHFULNESS and not for clarity, adequacy, and applicability. This refers to the correctness and includes accurate, adequate level of evidence, appropriate, believable, beyond doubt, credible, correct, convincing, factual, integral, objective, real, reasonable, right, trustful, unbiased, valid	3 TRUTHFULNESS How much are you satisfied with the following? (Scores: Full=2, Partial=1, No=0)	3.1 The title accurately describes the study 3.2 The variables investigated are valid to answer the research question 3.3 The design of the study is right to provide an unbiased answer to the research question (e.g., representative sample of the target population) 3.4 The subjects included are right to provide the correct answer to the research question 3.5 The sample size is correctly calculated for the stated reliability/power, and based on the correct variable or variables 3.6 Data collected and analyzed are factual, unbiased, and error free 3.7 Confounders and interaction duly accounted for 3.8 Correct method of analysis used considering the nature of the data (distribution, inter-dependence, adjustment for confounding, standardized rates, etc.)	

Contd...

Table 2: Contd...

Explanation of the domain	Domain	Item	Score or check
	3.9	Results really emanate from the data and analysis (not selected to serve a hypothesis, not manipulated), credible results - duly adjusted for missing values	
	3.10	Different results are internally consistent and internal and external validity of the results demonstrated	
	3.11	The interpretation of the results is valid in view of confounders, multiple <i>P</i> , etc.	
	3.12	Alternative explanations considered and ruled out	
	3.13	The conclusion is based on results, limitations, plausibility, corroborative evidence	
	Total score for TRUTHFULNESS (maximum 26)		
This section is only for APPLICABILITY and not for clarity, adequacy, or truthfulness. It includes accessible, affordable, beneficial, convenient, cost-effective, feasible, generalizable, harmless, impactful, important, interpretable, pragmatic, relevant, robust, scopeful, setting (community, hospital, clinic), simple, sustainable, timely, translationable, understandable, useful, valuable	4 APPLICABILITY		
	How much are you satisfied with the following? (Scores: Full=2, Partial=1, No=0)		
	4.1	The research question is relevant, useful, important, and timely	
	4.2	The setting of the study (clinic/hospital/community) is appropriate for application of the results, the target population is likely to benefit from the results, and inclusion-exclusion criteria are not too restrictive that will hamper the application	
	4.3	Variables used are such that the data on them can be collected in varying conditions:	
	4.4	Data elicitation methods and tools can be replicated by the others	
	4.5	Data used for results can be collected in a cost-effective manner	
	4.6	Analysis methods are understandable by any qualified professional, and replicable	
	4.7	The results are useful and applicable to the target population and robust/sustainable under varying conditions-replicability is demonstrated	
	4.8	Cost and convenience in implementing the results are reasonable and affordable, and the applicability and sustainability of the conclusion is convincing	
	4.9	The conclusion is medically important to change the current understanding, unaccounted uncertainties considered, and theoretical construct proposed	
Total score for APPLICABILITY (maximum 18)			
This section is for REPORT only comprising statement on all of the above aspects plus features such as articulate, brief, coherent, concise, focused, follow guidelines, logical, rganizer, succinct, articulate, simple, understandable,	5 REPORTING		
	How much are you satisfied with the following? (Scores: Full=2, Partial=1, No=0)		
	Description of all the previously listed items for which the scores are already assigned in sections 1 to 4 and not to be considered again here		
	5.1	Simple to understand and accessible to peers and non-specialists	
	5.2	The draft is concise, focused, and to the point (no unnecessary details)	
	5.3	Reporting guidelines (CONSORT/STROBE/STARD and SAMPL) followed	
	5.4	The presentation is in a logical sequence and coherent, including the use of the tables and graphs	
	5.5	Confidence but humility in presentation, keeping in view the limitations and medical uncertainties	
	Total score for REPORTING (maximum 10)		
	(Reporting of other items of quality already included in the other domains)		
	Total score (maximum 100)		

Note: These items in this table are stated for the draft of the paper. In case the research is at the planning or execution stage, the same items would apply to the concept and the process as anticipated or planned.

Such rigorousness is the price we should be prepared to pay for quality research. However, we have also proposed a QERM-Brief [Table 3], especially for the reviewers and editors, for relatively quick assessment although this also would require more time and effort of the reviewers than the present system. The journals may like to think of incentives to the reviewers such as priority consideration of their submission for publication. The reviewers may also consider the word limitation of the journal while assigning scores because this limit sometimes prevents giving full details. Now many journals accept supplementary material for their online version. This Brief version can be used by those researchers also who do not wish to be as rigorous with details.

This communication describes our efforts and the process for developing QERM scoring system for initiating a discussion on the need to develop such a system in the first place, to show that it is feasible, and to propose a credible scoring system to assess the quality of medical research for review, while the validation is in the process. Modifications as needed can be done.

Limitations

The proposed scoring system is only for empirical studies, namely descriptive studies, clinical trials, observational studies, and diagnostic studies. Other kinds of studies such as laboratory experiments, systematic reviews, and methodological research are excluded. The aspects such as plagiarism and data manipulation are not considered, presuming that all researchers

Table 3: Brief QERM score sheet for the reviewers (from the material as much as available from the manuscript) – to be used only for the studies that meet ethical standards

Item	Score: Satisfied=1; Not satisfied=0 (See explanation at the bottom)					Total
	Clarity ^a	Adequacy ^b	Truthfulness ^c	Applicability ^d	Reporting ^e	
1. The research question and the target population Worth investigating					X	
2. Abstract Properly summarizes the study					X	
3. Sample size and the subjects of the study Adequate sample size in consideration of reliability/power to detect a specified medically important effect, and representative of the population despite exclusions (dropout, outliers, etc.)					X	
4. Variables under study Antecedents, interventions, confounders, outcomes, rates, ratios, scores, etc.					X	
5. Study design Randomization and blinding in the case of clinical trials, selection of cases and controls, prospective/retrospective/cross-sectional design for observational studies, the sample design for descriptive studies					X	
6. Data elicitation tools and methods Case form, clinical assessments, laboratory results, interview method, instruments used, etc.					X	
7. Statistical analysis and interpretation Consideration of sample size, sampling method, repeated measures, correlations, confounding, interactions, etc.					X	
8. Results Restricted to the data available for the study, evidence-based, question answered					X	
9. Discussion Includes review of the current literature, interpretation of the results, resolution of conflicts with other findings, corroborative evidence, biological plausibility, and limitations					X	
10. Conclusion Combination of results, plausibility, corroborative evidence, and limitations – should be convincing (face-valid) and applicable					X	
11. Presentation (out of 10) Concise, coherent, unambiguous, clear, and understandable, supported by tables and graphs, and follow the reporting guidelines.						
Total (out of a maximum of 50)						

^aClarity includes fully identified, open, transparent, unambiguous, well defined. ^bAdequacy includes comprehensive, complete, enough, ethical, focused, full, justified, measurable, novel, original, plausible, rational, reliable, repeatable, replicable, reproducible, rigorous, significant, sufficient, worthy. ^cTruthfulness includes accurate, an adequate level of evidence, appropriate, believable, beyond doubt, credible, correct, convincing, factual, integral, objective, real, reasonable, right, trustful, unbiased, valid. ^dApplicability includes accessible, affordable, beneficial, convenient, cost-effective, feasible, generalizable, harmless, impactful, important, interpretable, pragmatic, relevant, robust, scopeful, setting (community, hospital, clinic), simple, sustainable, timely, translationable, understandable, useful, valuable. ^eReporting includes a statement of all of the above: Articulate, brief, coherent, complete, concise, focused, follow guidelines, logical, organized, succinct, understandable. X Reporting of these items is already included in the other four domains

recognize these as malpractices. The scoring systems are verified for the face and content validity and the other kinds of validations are still in process.

Conclusion

Research resources are scarce, and they must be spent optimally for beneficial outcomes. Assessing the quality of medical research starting from the planning stage can

substantially help in improving the quality—thus in achieving the outcomes that really contribute to improved health. A high QERM score may require meticulous research process from conceiving the research question to drafting the paper for publication. We believe that awareness of such a scoring system and its usage can substantially improve the quality of medical research. At the same time, this may make the research process much more difficult, but quality comes at a cost.

Our scoring system is verified for face validity and content validity as assessed by the experts in the Group. It certainly needs additional evidence which would come from validation and wider applications in diverse settings.

Authors' contribution

AI proposed the idea and piloted the project. All significantly contributed to the deliberations and the draft. All approved the manuscript.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Ioannidis JP. How to make more published research true. *PloS Med* 2014;11:e1001747.
- Mendoza D, Garcia CA. Defining research reproducibility: What do you mean? *Clin Chem* 2017;63:1777.
- ESHRE Capri Workshop Group. Protect us from poor-quality medical research. *Hum Reprod* 2018;33:770–6.
- Mische SM, Fisher NC, Meyn SM, Sol-Church K, Hegstad-Davies RL, Weis-Garcia F, *et al.* A review of the scientific rigor, reproducibility, and transparency studies conducted by the ABRF Research Groups. *J Biomol Tech* 2020;31:11–26.
- Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283–4.
- Ioannidis JP. Why most published research findings are false. *PloS Med* 2005;2:e124.
- Shwartz M, Restuccia JD, Rosen AK. Composite measures of health care provider performance: A description of approaches. *Milbank Q* 2015;93:788–825.
- Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
- Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
- Kane R, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 2007;60:241–9.
- Hendriksma M, Joosten MH, Peters JP, Grolman W, Stegeman I. Evaluation of the quality of reporting of observational studies in otorhinolaryngology – based on the STROBE statement. *PLoS ONE* 2017;12:e0169316.
- Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, *et al.* Reporting diagnostic accuracy studies: Some improvements after 10 years of STARD. *Radiology* 2015;274:781–9.
- Chambers C. The registered reports revolution: Lessons in cultural reform. *Significance* 2019;16:23–7.
- PloS ONE Guidelines for Reviewers. Available from <https://journals.plos.org/plosone/s/reviewer-guidelines>. [Last accessed on 2022 Jun 16].
- Wiley. Journal Reviewers 2020. Available from <https://authorservices.wiley.com/Reviewers/journal-reviewers/index.html>. [Last accessed on 2022 Jun 16].
- 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;276:637–9.
- Fernández E, STROBE group. Estudios epidemiológicos (STROBE) [Observational studies in epidemiology (STROBE)]. *Med Clin (Barc)* 2005;125(Suppl 1):43–8.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003;49:1–6.
- Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, *et al.* The SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines for quality improvement reporting: Explanation and elaboration. *Qual Saf Health Care* 2008;17(Suppl 1):i13–32.
- ICMJE. Preparing a Manuscript for Submission to a Medical Journal. Available from: <http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html>. [Last accessed on 2022 Jun 16].
- Oxman A, Guyatt G. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44:1271–8.
- Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, *et al.* From systematic reviews to clinical recommendations for evidence-based health care: Validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J* 2010;4:84–91.
- Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- Newcastle Ottawa Scale. Manual. Available from: http://www.ohri.ca/programs/clinical_epidemiology/nos_manual.pdf. [Last accessed on 2022 Jun 16].
- Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–7.
- Mays N, Pope C. Qualitative research in health care: Assessing quality in qualitative research. *BMJ* 2000;320:50–2.
- Ten Have H, Gordijn B. Medical epistemology. *Med Health Care Philos* 2017;20:451–2.
- Kemper JM, Wang HTY, Ong AGJ, Mol BW, Rolnik DL. The quality and utility of research in ectopic pregnancy in the last three decades: An analysis of the published literature. *Eur J Obstet Gynecol Reprod Biol* 2020;245:134–42.
- Rajabally YA, Fatehi F. Outcome measures for chronic inflammatory demyelinating polyneuropathy in research: Relevance and applicability to clinical practice. *Neurodegener Dis Manag* 2019;9:259–66.
- Scott-Findlay S, Pollock C. Evidence, research, knowledge: A call for conceptual clarity. *Worldviews Evid Based Nurs* 2004;1:92–7.
- Shaw D. The quest for clarity in research integrity: A conceptual schema. *Sci Eng Ethics* 2019;25:1085–93.
- Montenegro-Montero A, García-Basteiro AL. Transparency and reproducibility: A step forward. *Health Sci Rep* 2019;2:e117.
- Altman DG, Moher D. Declaration of transparency for each research article. *BMJ* 2013;347:f4796.
- Davidoff F. News from the international committee of medical journal editors. *Ann Intern Med* 2000;133:229–31.
- Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, *et al.* Scientific standards. Promoting an open research culture. *Science* 2015;348:1422–5.
- Bordage G. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. *Acad Med* 2001;76:889–96.
- Pierson DJ. The top 10 reasons why manuscripts are not accepted for publication. *Respir Care* 2004;49:1246–52.
- Begley CG, Ioannidis JP. Reproducibility in science: Improving the standard for basic and preclinical research. *Circ Res* 2015;116:116–26.
- Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, *et al.* A manifesto for reproducible science. *Nature Hum Behav* 2017.
- Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature. *PLoS Biol* 2018;16:e2006930.
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612–3.
- Plesser HE. Reproducibility vs. replicability: A brief history of a confused terminology. *Front Neuroinform* 2018;11:76.
- Bollen K, Cacioppo JT, Kaplan RM, Krosnick JA, Olds JL, Dean H. Social, behavioral, and economic sciences perspectives on robust and reliable science. National Science Foundation 2015. P3. Available from: https://nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf. [Last accessed on 2022 Jun 16].
- Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med* 2016;8:341ps12.
- Indrayan A, Mishra A. The importance of small samples in medical

- research. *J Postgrad Med* 2021;67:219–23.
46. Indrayan A, Holt M. *Concise Encyclopedia of Biostatistics for Medical Professionals*. Boca Raton, CRC Press; 2017.
47. Holme C. Cultivate absolutely accuracy in observation and truthfulness in report. *J Adv Nurs* 2020;76:1093–4.
48. Andrade C. HARKing, Cherry-picking, P-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *J Clin Psychiatry* 2021;82:20f13804.
49. Tarran B. New year, familiar problems. *Significance* 2020;17:1.
50. Booth A, Brice A. *Evidence-based Practice for Information Professionals: A Handbook*. Vol. 21. London Facet Publishing; 2004.
51. Ioannidis JP. Why most clinical research is not useful. *PloS Med* 2016;13:e1002049.
52. Knottnerus JA, Tugwell P. How to write a research paper. *J Clin Epidemiol* 2013;66:353–4.
53. Cooper ID. How to write an original research paper (and get it published). *J Med Libr Assoc* 2015;103:67–8.
54. Zieger M. *Essentials of Writing Biomedical Research Papers*. 2nd ed. New York, McGraw Hill; 1999.
55. Indrayan A. Reporting of basic statistical methods in biomedical journals: Improved SAMPL guidelines. *Indian Pediatr* 2020;57:43–8.
56. Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures? *Med Care* 2005;43:1177–84.
57. Sandercock P, Whiteley W. How to do high-quality clinical research 1: First steps. *Int J Stroke* 2018;13:121–8.
58. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, *et al.* Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials* 1996;17:1–12.
59. Catillon M. Trends and predictors of biomedical research quality, 1990–2015: A meta-research study. *BMJ Open* 2019;9:e030342.
60. Higgins J, Altman D, Gøtzsche P, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in 24 randomised trials. *BMJ* 2011;343:d5928.
61. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: A systematic review. *Phys Ther* 2008;88:156–75.
62. Gabriel A, Maxwell GP. Reading between the lines: A plastic surgeon's guide to evaluating the quality of evidence in research publications. *Plast Reconstr Surg-Glob Open* 2019;7:e2311.
63. Indrayan A, Malhotra RK. *Medical Biostatistics*. 4th ed. Boca Raton Chapman & Hall/CRC Press; 2018.
64. Montagna E, Zaia V, Laporta GZ. Adoption of protocols to improve quality of medical research. *Einstein (São Paulo)* 2020;18:1–4.
65. Ueda R, Nishizaki Y, Homma Y, Sanada S, Otsuka T, Yasuno S, *et al.* Importance of quality assessment in clinical research in Japan. *Front Pharmacol* 2019;10:1228.
66. Glasziou P, Vandenbroucke J, Chalmer I. Assessing the quality of research. *BMJ* 2004;328:39–41.

The Supplementary Material is available online

<https://www.jpgmonline.com/article.asp?issn=0022-3859;year=2022;volume=68;issue=4;spage=221;epage=230;aualast=Indrayan>