


PERSPECTIVE

Reporting guideline checklists are not quality evaluation forms: they are guidance for writing

Patricia Logullo¹  | Angela MacCarthy¹ | Shona Kirtley¹ | Gary S. Collins^{1,2}

¹UK EQUATOR Centre, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics & Musculoskeletal Sciences, University of Oxford, Oxford, UK

²NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

Correspondence

Dr. Patricia Logullo, UK EQUATOR Centre, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics & Musculoskeletal Sciences, University of Oxford, Botnar Research Centre - Old Road, Oxford OX3 7LD, UK.
Email: patricia.logullo@ndorms.ox.ac.uk

KEYWORDS: 4-6, MesH

One of the fundamental principles of health research integrity is that research methods and results should be completely and transparently reported. Clear, detailed reporting allows the reader to understand how a study was designed and conducted, to judge the reliability of its findings and the reproducibility of its methods, and to use the tested interventions in their clinical practice.¹⁻³ The way in which research results are reported, therefore, can have a direct impact on patients' lives.⁴ As the late Professor Douglas Altman said, '*Readers should not have to infer what was probably done, they should be told explicitly*'.⁵

Reporting guidelines were created to help researchers write reports that contain the minimum set of information necessary to allow readers to clearly understand what was done and found in a study and facilitate a formal risk of bias assessment (using tools such as the Cochrane Risk of Bias tool or QUADAS). Complete reporting can also allow replication of study methods and procedures. 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*'.⁶ Following the publication of the first reporting guideline for clinical trials, CONSORT, in 1996,⁷ multiple reporting guidelines have been published, covering a range of study designs (eg, clinical trials, observational studies), clinical areas (eg, nutrition), or parts of a report (eg, abstracts), to help biomedical researchers write up their studies for publication.^{8,9} Stakeholders in biomedical research have embraced reporting guidelines, with major funders and a large number of biomedical journals endorsing the guidelines and increasingly requiring their use.^{10,11}

The most widely used and well-known reporting guidelines usually consist of a statement paper that describes the process of

developing the guideline and presents the guideline usually in the form of a 'checklist'.⁴ Each checklist consists of a different number of reporting content items, ranging from just a few to more than 30 items. These checklists are designed to be easy to use by authors when they start writing their manuscript. Many journals have recognised how useful they are and have implemented reporting guidelines in their submission and editorial processes. Several journals also require authors to submit a completed checklist indicating where in the manuscript each item has been reported.

Reporting guidelines are (or at least should be) rigorously developed following an extensive process of expert consultation and should not reflect just the opinion of one individual⁶; they should represent a consensus-based minimal set of items that a group of experienced researchers, journal editors, policymakers, and other stakeholders (eg, funders, patient representatives) have determined should be reported.

WHAT IS THE OUTCOME BEING MEASURED?

Whilst designed to help improve the completeness and transparency of reporting, reporting guidelines are increasingly used to determine the 'quality' of a research paper. However, there are many problems with this. One major issue relates to the concept of quality itself. While some researchers might think that a 100% adherence to a set of content reporting items would mean 'a quality paper', others might argue that this 'top quality' is not attainable and manuscripts adhering to, say, 80% of the items are 'well reported'. Therefore, there should first be a consensus—ideally agreed by reporting guideline authors—

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Health Science Reports* published by Wiley Periodicals LLC

about determining what level of quality is needed for a health research article to be considered 'well reported'; in other words, define what quality of reporting is. This is, however, what properly developed reporting guidelines do: they outline a minimum set of information that should be reported in health research manuscripts. This minimum set of information items compose and define a 'total quality' report, and researchers should ensure that they indeed describe every item in their manuscripts.

However, if one defines 'reporting quality' as 100% adherence to a reporting checklist, understood as the adherence to all items of a given reporting guideline, then it will be virtually impossible to find a 'good report' in currently published research. On the other hand, if the outcome is too broadly defined and not standardized, such flexibility might put two very different papers under the same category of 'good report'. For example, the same manuscript may be evaluated as a 'good report' by a study considering 70% of adherence to a reporting guideline, while another study would find this same manuscript not so good because the authors expected 80% to be a minimum adherence indicating quality. Similarly, manuscripts may have the same level of adherence but cover different aspects of the reporting guideline, as different researchers can consider different items as key or ancillary. 'Reporting quality', therefore, is a very subjective concept. Published studies do not agree on how much quality to expect—and maybe they should all expect 100% adherence as per the definition of reporting guidelines: a **minimum** set of information.

QUALITY EVALUATION TOOLS?

Numerous studies have now been published evaluating whether individual reporting guidelines have made any improvement to the completeness of published reports.^{12–14} These studies typically use adherence to a reporting guideline as a surrogate for reporting quality^{15–41} or even, inadequately, for study quality.⁴² The findings of such research-on-research studies generally agree that the quality of health research reports is still lacking.⁴³ However, the methods used to investigate this complex concept of 'quality of publication' varies widely in the literature. In most cases, the original reporting guideline checklist is being used without modification to measure 'quality'—which is a complex concept on its own—but there is no consensus on whether or how to apply these reporting guidelines in studies on adherence.

One might argue that because reporting guidelines are the result of carefully planned discussions at consensus meetings, their face validity would be guaranteed, in the sense that all items in the checklist are considered relevant or essential. However, that does not mean that when experts develop reporting checklists, they do so with the intention that the checklist will also serve as a properly designed evaluation tool for assessing reporting quality; reporting guidelines are specifically designed as guidance for writing. The STREGA reporting guideline explicitly indicates this: '*the STREGA reporting guidelines should not be used for screening submitted manuscripts to determine the quality or validity of the study being reported*'.⁴⁴

One exception in the literature, however, is the TRIPOD guideline.^{45–47} The TRIPOD Statement is a reporting guideline for prediction models (TRIPOD stands for *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*).^{45–47} TRIPOD authors, recognising the widespread secondary use of reporting guidelines, set out to develop and publish an evaluation form for assessing the quality of reporting of diagnostic and prognostic prediction model studies. This form can be used by any researcher trying to evaluate the quality of prediction models in the literature, facilitating the comparison of results of different studies (Table 1).^{47,48}

Table 1 shows an example of one checklist item (item 4) from the TRIPOD reporting guideline. The exact text from the TRIPOD reporting checklist is contained in column 1. Column 2 provides the text from the TRIPOD evaluation tool, which breaks down the item into several questions. Columns 3 to 6 provide information about how to score the reporting of item 4. The Table shows that in order to conduct a robust evaluation of the reporting of checklist items, simply relying on the reporting checklist items themselves is not enough. Each item needs to be broken down into appropriate questions, with an accompanying scoring system developed. Building such an evaluation tool for each reporting guideline will enable researchers to consistently scrutinise and score the reporting quality of research papers, with every researcher around the world using the same tool, as it happens with quality of life evaluations, for example, an outcome that can be compared among studies when they use the same tool.^{49,50}

SCORING SYSTEMS

Another important issue is the design and content of the data extraction form used to evaluate 'reporting quality' in these studies. How do researchers assign a score to each reporting checklist item in these evaluation forms? Currently, there seems to be no consistency in the methods or scoring systems being used by researchers.^{15–40} Some studies evaluate simply whether an item is reported or not (a 'yes/no' dichotomised score).^{19,25,29} Others assign three options, for example, 'not reported', 'fully reported', and 'partially reported' or 'not applicable'.^{15,17,20–24,26,27,31,33,37–40} Some studies also use more options, such as a five-point scale of quality for each item.^{28,32,35} Given the variability in scoring adherence between studies (ie, each study gives different weights to the same item), how can the results of these studies be compared?

One might propose that it is sufficient to include a 'not applicable' option to the reporting guideline checklist items when developing a scoring system, and it would be ready to use as an evaluation tool. But this may not be enough. The authors of TRIPOD discuss:

Overall adherence, in the form of a percentage of items adhered to, requires a clear denominator of total number of items one can adhere to. One has to decide whether to take items that are considered not applicable into account in the numerator as well as in the denominator. Determining applicability is subjective

TABLE 1 Example of checklist items turned into evaluation form questions in the TRIPOD reporting guideline, for prediction models for prognosis or diagnosis

| Item | Original reporting guideline checklist item | # | Evaluation form items | Instructions for scoring | | | |
|------|---|-----|--|---------------------------------------|---------------------------------------|--|--|
| | | | | D Score 1 if element is scored as 'Y' | V Score 1 if element is scored as 'Y' | IV Score 1 if element is scored as 'Y' | D + V Score 1 if element is scored as 'Y' |
| | | | | | | | |
| 4a | 'Describe the study design or source of data (eg, randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable'. | i | <p>The study design/source of data is described</p> <p>For example, Prospectively designed, existing cohort, existing RCT, registry/medical records, case control, case series.</p> <p>This needs to be explicitly reported; reference to this information in another article alone is insufficient.</p> | Y/N | Y/N | Y/N | =Y if D4ai = Y AND V4ai = Y |
| 4b | 'Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up'. | i | The starting date of accrual is reported | Y/N/R | Y/N/R | Y/N/R | =Y if (D4bi = Y AND V4bi = [Y OR R]) OR (D4bi = [Y OR R] AND V4bi = Y) = R if D4bi = R AND V4bi = R |
| | | ii | The end date of accrual is reported | Y/N/R | Y/N/R | Y/N/R | =Y if (D4bii = Y AND V4bii = [Y OR R]) OR (D4bii = [Y OR R] AND V4bii = Y) = R if D4bii = R AND V4bii = R |
| | | iii | <p>The length of follow-up and prediction horizon/time frame are reported, if applicable</p> <p>E.g. 'Patients were followed from baseline for 10 years' and '10-year prediction of...'; notably for prognostic studies with long term follow-up.</p> <p>If this is not applicable for an article (ie, diagnostic study or no follow-up), then score Not applicable.</p> | Y/N/NA | Y/N/NA | Y/N/NA | =Y if (D4biii = Y AND V4biii = [Y OR NA]) OR (D4biii = [Y OR NA] AND V4biii = Y) = NA if D4biii = NA AND V4biii = NA |
| | | | | | | | |

Abbreviations: Y, yes; N, no; N/A, not applicable; R, referenced; D, development (applies for studies that develop new prediction models); V, external validation (applies for studies that validate existing models); IV, applies for studies of incremental value; D + V, applies for studies of development and external validation of the same model.

and requires interpretation. In our experience, items for which interpretation was needed, sometimes indicated by phrases like 'if relevant' or 'if applicable,' were the most difficult ones to score and these items are a potential threat to inter-assessor agreement.

As the number of papers assessing the quality of reporting of studies is increasing, it is important to highlight the pitfalls of using reporting guideline checklists as evaluation tools. It seems that the only way to prevent multiple methodologists from assessing manuscript quality using different criteria, forms, scoring systems, outcomes, and number of evaluators is to provide clear guidance on how to evaluate the reporting quality of manuscripts and to encourage all reporting guideline developers to publish a reporting evaluation tool together with or soon after the publication of a new reporting guideline. Providing an evaluation form would, at least, offer evaluators a single tool to be used uniformly across studies, allowing some comparability.

DEVELOPMENT AND TESTING OF EVALUATION TOOLS

There are several methodological steps that researchers must follow when developing evaluation tools to ensure the relevance and robustness of a new tool to evaluate a subjective concept, for instance, quality of life. An evaluation instrument such as a questionnaire or scoring system (ie, composed of multiple parts or items, taken as indirect indicators) must undergo validity testing before it can be said to accurately measure what it intends to measure, that it is clear and easily understandable for users, and that it represents all facets of a (sometimes complex) concept. Where other instruments exist, it is possible to validate the results of a new tool by comparing it to the other, considered, so far, a 'gold standard'. It is desirable that the instrument has some consistency over time too, measuring the same thing the same way twice, or by different evaluators.

As far as we know, none of these methods traditionally used in health outcome measurement have been followed when developing reporting guideline checklists. Perhaps this is because reporting quality is seen as an objective outcome: the 100% adherence to a checklist. Perhaps it is because the developers did not set out to develop an evaluation tool in the first place, but only guidance for writing, the exception being the TRIPOD evaluation tool, mentioned earlier, which was developed in addition to the reporting guideline checklist.

There are currently at least 84 reporting guidelines under construction, according to the EQUATOR Network registry (<https://www.equator-network.org/library/reporting-guidelines-under-development/>); more, if we consider that not every development team registers their guideline under development. Developers should consider building evaluation tools along with their reporting guideline. However, when this is not possible (eg, due to lack of funding), they should follow the example of the STREGA authors⁵¹ and warn researchers not to use their reporting guideline as a quality evaluation

tool. Existing reporting guideline groups should also be encouraged to develop evaluation tools for their guidelines. This will ensure that, in the future, all research studies assessing adherence to reporting guidelines or measuring the 'quality' of reporting will use robustly and appropriately developed evaluation tools, and the results will be more meaningful and reliable.

ACKNOWLEDGEMENTS

P.L., A.M., S.K. and G.S.C. are funded by Cancer Research UK (programme grant C49297/A27294). GSC was supported by the NIHR Biomedical Research Centre, Oxford.

AUTHOR CONTRIBUTIONS

Conceptualization: Patricia Logullo, Gary S. Collins

Data Curation: Patricia Logullo, Angela MacCarthy, Gary S. Collins

Formal Analysis: Patricia Logullo, Gary S. Collins

Funding Acquisition: Gary S. Collins

Resources: Gary S. Collins

Writing - Original Draft: Patricia Logullo, Shona Kirtley, Gary S. Collins

Writing - Review & Editing: Angela MacCarthy, Shona Kirtley, Gary S. Collins

All authors have read and approved the final version of the manuscript.

CONFLICT OF INTEREST

Gary Collins is involved in the TRIPOD Statement.

ORCID

Patricia Logullo  <https://orcid.org/0000-0001-8708-7003>

REFERENCES

1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374:86-89.
2. Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383(9913):257-266.
3. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
4. MacCarthy A, Kirtley S, de Beyer JA, Altman DG, Simera I. Reporting guidelines for oncology research: helping to maximise the impact of your research. *Br J Cancer*. 2018;118(5):619-628.
5. Altman D. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ*. 1996;313:570-571.
6. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.
7. Begg CC, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials - the CONSORT statement. *JAMA*. 1996;276(8):637-639.
8. Moher D, Altman DG. Four proposals to help improve the medical research literature. *PLoS Med*. 2015;12(9):e1001864.
9. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR network. *J R Soc Med*. 2016;109(2):67-77.
10. Research NCfRRRoAi. ARRIVE at five 2016. <https://www.nc3rs.org.uk/news/arrive-five>.

11. Endorsers - Journals and Organizations 2018. <http://www.consort-statement.org/about-consort/endorsers1>.
12. Jin Y, Sanger N, Shams I, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? - a systematic review of reviews: an update. *J Multidiscip Healthc*. 2018;11:495-510.
13. Li G, Mbuagbaw L, Samaan Z, et al. State of reporting of primary biomedical research: a scoping review protocol. *BMJ Open*. 2017;7(3):e014749.
14. Li G, Abbade LPF, Nwosu I, et al. A scoping review of comparisons between abstracts and full reports in primary biomedical research. *BMC Med Res Methodol*. 2017;17(1):181.
15. Limaye D, Limaye V, Pitani RS, et al. Development of a quantitative scoring method for Strobe checklist. *Acta Pol Pharm Drug Res*. 2018;75(5):1095-1106.
16. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open*. 2019;9(5):e026589.
17. Blanco D, Biggane AM, Cobo E, MiRo Rn. Are CONSORT checklists submitted by authors adequately reflecting what information is actually reported in published papers? *Trials*. 2018;19(1):80.
18. Botos J. Reported use of reporting guidelines among JNCI: journal of the National Cancer Institute authors, editorial outcomes, and reviewer ratings related to adherence to guidelines and clarity of presentation. *Res Integr Peer Rev*. 2018;3:7.
19. Chantebel R, Chesneau A, Tavernier E, El-Hage W, Caille A. Completeness of descriptions of repetitive transcranial magnetic stimulation intervention: a systematic review of randomized controlled trials of rTMS in depression. *J ECT*. 2019;35(1):7-13.
20. Chow JTY, Turkstra TP, Yim E, Jones PM. The degree of adherence to CONSORT reporting guidelines for the abstracts of randomised clinical trials published in anaesthesia journals: a cross-sectional study of reporting adherence in 2010 and 2016. *Eur J Anaesthesiol*. 2018;35:942-948.
21. Godinho MA, Gudi N, Milkowska M, Murthy S, Bailey A, Nair NS. Completeness of reporting in Indian qualitative public health research: a systematic review of 20 years of literature. *J Public Health (Oxf)*. 2019;41(2):405-411.
22. Goi PD, Goi JD, Cordini KL, Cereser KM, Rocha NS. Evaluating psychiatric case-control studies using the STROBE (Strengthening the reporting of OBservational studies in epidemiology) statement. *Sao Paulo Med J*. 2014;132(3):178-183.
23. Hair K, Macleod MR, Sena ES. A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *bioRxiv*. 2018.
24. Han S, Olonisakin TF, Pribis JP, et al. A checklist is associated with increased quality of reporting preclinical biomedical research: a systematic review. *PLoS One*. 2017;12(9):e0183591.
25. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.
26. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVED: support for the ARRIVE (animal research: reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS One*. 2018;13(5):e0197882.
27. Ramke J, Palagyi A, Jordan V, Petkovic J, Gilbert CE. Using the STROBE statement to assess reporting in blindness prevalence surveys in low and middle income countries. *PLoS One*. 2017;12(5):e0176178.
28. Rao A, Bruck K, Methven S, et al. Quality of reporting and study design of CKD cohort studies assessing mortality in the elderly before and after STROBE: a systematic review. *PLoS One*. 2016;11(5):e0155078.
29. Rocchietta I, Nisand D. A review assessing the quality of reporting of risk factor research in implant dentistry using smoking, diabetes and periodontitis and implant loss as an outcome: critical aspects in design and outcome assessment. *J Clin Periodontol*. 2012;39(Suppl 12):114-121.
30. Samaan Z. A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc*. 2013;6:169-188.
31. Saric L, Vucic K, Dragicevic K, et al. Comparison of conference abstracts and full-text publications of randomized controlled trials presented at four consecutive world congresses of pain: reporting quality and agreement of results. *Eur J Pain*. 2019;23(1):107-116.
32. Serrano M, Gonzalvo MC, Sanchez-Pozo MC, et al. Adherence to reporting guidelines in observational studies concerning exposure to persistent organic pollutants and effects on semen parameters. *Hum Reprod*. 2014;29(6):1122-1133.
33. Sorensen AA, Wojahn RD, Manske MC, Calfee RP. Using the Strengthening the reporting of observational studies in epidemiology (STROBE) statement to assess reporting of observational trials in hand surgery. *J Hand Surg Am*. 2013;38(8):1584-1589. e2.
34. Stevens A, Shamseer L, Weinstein E, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ*. 2014;348:g3804.
35. Svenkerud S, MacPherson H. The impact of STRICTA and CONSORT on reporting of randomised control trials of acupuncture: a systematic methodological evaluation. *Acupunct Med*. 2018;36(6):349-357.
36. Turner L, Larissa S, Altman DG, Schulz KF, Moher D. Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? *Cochrane Rev Syst Rev*. 2012;2:60.
37. Song JW, Guiry SC, Shou H, et al. Qualitative assessment and reporting quality of intracranial vessel wall MR imaging studies: a systematic review. *AJNR Am J Neuroradiol*. 2019;40(12):2025-2032.
38. Adams AD, Benner RS, Riggs TW, Chescheir NC. Use of the STROBE checklist to evaluate the reporting quality of observational research in obstetrics. *Obstet Gynecol*. 2018;132(2):507-512.
39. Aghazadeh-Attari J, Mobaraki K, Ahmadzadeh J, Mansorian B, Mohebbi I. Quality of observational studies in prestigious journals of occupational medicine and health based on Strengthening the reporting of observational studies in epidemiology (STROBE) statement: a cross-sectional study. *BMC Res Notes*. 2018;11(1):266.
40. Madden K, Phillips M, Solow M, McKinnon V, Bhandari M. A systematic review of quality of reporting in registered intimate partner violence studies: where can we improve? *J Inj Violence Res*. 2019;11(2):123-136.
41. da Costa BR, Cevallos M, Altman DG, Rutjes AW, Egger M. Uses and misuses of the STROBE statement: bibliographic study. *BMJ Open*. 2011;1(1):e000048.
42. Caulley L, Catalá-López F, Whelan J, et al. Reporting guidelines of health research studies are frequently used inappropriately. *J Clin Epidemiol*. 2020;122:87-94.
43. Glasziou P, Chalmers I. Research waste is still a scandal-an essay by Paul Glasziou and Iain Chalmers. *BMJ*. 2018;363:k4645.
44. Goodman D, Ogrinc G, Davies L, et al. Explanation and elaboration of the SQUIRE (standards for quality improvement reporting excellence) guidelines, V.2.0: examples of SQUIRE elements in the healthcare improvement literature. *BMJ Qual Saf*. 2016;25(12):e7.
45. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Cancer*. 2015;112(2):251-259.
46. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W73.
47. Heus P, Damen J, Pajouheshnia R, et al. Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for

- assessing completeness of reporting of prediction model studies. *BMJ Open*. 2019;9(4):e025611.
48. Assessing adherence of prediction model reports to the TRIPOD guideline 2018. https://www.tripod-statement.org/Portals/0/Documents/Downloads/TRIPOD%20Adherence%20assessment%20form_V-2018_12.pdf.
49. Wells GA, Russell AS, Haraoui B, Bissonnette R, Ware CF. Validity of quality of life measurement tools-from generic to disease-specific. *J Rheumatol Suppl*. 2011;88:2-6.
50. Lorente S, Vives J, Viladrich C, Losilla JM. Tools to assess the measurement properties of quality of life instruments: a meta-review protocol. *BMJ Open*. 2018;8(7):e022829.
51. Little J, Higgins JP, Ioannidis JP, et al. Strengthening the REporting of genetic association studies (STREGA): an extension of the STROBE statement. *PLoS Med*. 2009;6(2):e22.

How to cite this article: Logullo P, MacCarthy A, Kirtley S, Collins GS. Reporting guideline checklists are not quality evaluation forms: they are guidance for writing. *Health Sci Rep*. 2020;09:e165. <https://doi.org/10.1002/hsr2.165>