Guidance on Statistical Reporting to Help Improve Your Chances of a Favorable Statistical Review

Over our tenure as the statistical editors of AJRCCM and AnnalsATS, we have observed recurrent methodological issues and reporting practices in submitted manuscripts that invariably lead to unfavorable reviews by statistical reviewers, content reviewers, and editorial board members. In an effort to help authors improve both the statistical rigor and clinical impact of their manuscripts, we have developed this document to both combine our suggestions and centralize resources and references that authors can use to avoid common pitfalls and improve reporting quality.

Clearly State the Aims of the Study in the Introduction

To assess whether the selected methods are appropriate for a study, the goals and specific hypotheses being tested must be clearly stated. This is often not the case. We recommend that authors use a PICO (population, intervention, comparator, outcome) (1) template or extension (2) (tailored PICO templates are available for most study designs) to develop clear study aims and research questions.

Follow Relevant Reporting Guidelines and Provide Details Regarding Analytic Decisions

Providing a full accounting of the study design, data collection, and data analysis can seem like an overwhelming task in any study, particularly within allowed word count limits. Thankfully, although a study question might be novel, the study design is usually not, and authors should take advantage of the many available guidelines and checklists that have been developed to detail what information should be reported for a given study design. To promote awareness, in Table 1, we list the guidelines for many common study designs, all of which are available from the EQUATOR (Enhancing the Quality and Transparency of Health Research) Network (www.equator-network.org/). The easiest way for authors to enhance the quality of their manuscript is to include all of the items and elements listed in the appropriate reporting guideline and accompanying checklist for their specific study design. These guidelines also provide a natural structure and sequence for authors to follow when writing their manuscript, because reporting elements are usually separated by each section of a manuscript.
Table 1. Reporting Guidelines for Common Study Designs

| Study Type                        | Reporting Guidelines |
|----------------------------------|----------------------|
| Randomized trial                 | CONSORT              |
| Observational study*             | STROBE               |
| Systematic review                 | PRISMA               |
| Study protocol                    | SPIRIT, PRISMA-P     |
| Prediction model study†           | TRIPOD               |
| Case report                       | CARE                 |
| Clinical practice guideline       | AGREE, RIGHT         |
| Qualitative research              | SRQR, COREQ          |
| Animal preclinical study          | ARRIVE               |
| Quality improvement study         | SQUIRE               |
| Economic evaluation               | CHEERS               |

Definition of abbreviations: AGREE = Appraisal of Guidelines, Research and Evaluation; ARRIVE = Animal Research: Reporting of In Vivo Experiments; CARE = Case Report guidelines (www.care-statement.org); CHEERS = Consolidated Health Economic Evaluation Reporting Standards; CONSORT = Consolidated Standards of Reporting Trials; COREQ = Consolidated Criteria for Reporting Qualitative Studies; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA-P = Preferred Reporting Items for Systematic Review and Meta-Analyses Protocols; RIGHT = Reporting Items for Practice Guidelines in Healthcare; SPIRIT = Standard Protocol Items: Recommendations for Interventional Trials; SQUIRE = Standards for Quality Improvement Reporting Excellence; SRQR = Standards for Reporting Qualitative Research; STROBE = Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD = Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

At the time of publication, a total of 424 reporting guidelines were available at the EQUATOR Network website (www.equator-network.org), covering almost all potential research activities. Adapted by permission from Reference 20.

See also recent guidance for observational causal inference studies in pulmonary, critical care, and sleep medicine (9).

We are also strong proponents of the SAMPL (Statistical Analyses and Methods in the Published Literature) guidelines (4). The SAMPL guidelines are applicable to almost any statistical analysis or numeric value that authors may conduct or report, respectively. We also recommend that authors explore the statistical analysis and reporting guidance provided by Assel and colleagues (5) and the Annals of Internal Medicine (6), both of which are open access.

Though these are mostly requested in the aforementioned study-specific guidelines, we also wish to highlight a few persistently problematic or omitted areas of manuscripts (Table 2). For all studies, authors must report any missing data and then discuss and justify their approach for handling missing data (e.g., imputation method and variables used). In observational studies in which researchers are seeking to describe a relationship or association between an exposure and an outcome, we want to see a rationale for all included confounder variables and a discussion (and potentially an analysis [7, 8]) of the potential impact of missing confounders. For additional guidance, we recommend the causal inference resource by Lederer and colleagues (9) and that authors follow the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement (10) and checklist (Table 1).

Table 2. Key Elements That Should Be Reported in Research Manuscripts

- Indicate that the appropriate reporting guideline (Table 1) was followed. We encourage authors to include the accompanying checklist, if there is one for the guidelines they are using, in the supplementary material.
- Include a reference number for the ethics board or animal care and use approval. This is the legal authority for the work and allows traceability for the use of data and human samples.
- Report information regarding preregistration of clinical trials and systematic reviews. Any changes or deviations from original and published protocols, particularly for trials, should be reported. We check www.clinicaltrials.gov, PROSPERO, and other repositories. We are increasingly confirming that results of clinical trials have been submitted to www.clinicaltrials.gov or EudraCT within the required time frame (i.e., within 1 yr of study completion for www.clinicaltrials.gov).
- Report when and where the study was completed.
- Report the potentially eligible and final sample sizes, together with study inclusion and exclusion criteria, ideally in a flowchart. Also explain the rationale for choosing the inclusion and exclusion criteria.
- Include a detailed description and rationale for the statistical analyses. This description should include the testing of assumptions associated with the selected statistical model(s), how missing data were handled, any sensitivity analyses, and information regarding the statistical software and packages used by the authors.

Definition of abbreviations: EudraCT = European Union Drug Regulating Authorities Clinical Trials Database; PROSPERO = international prospective register of systematic reviews in health and social care. Because some journals (e.g., AJRCCM and AnnalsATS) favor a brief description of the methods in the main manuscript, authors must achieve a balance between what is reported in the main manuscript text and the online supplement. We suggest that authors focus on key data and analytic details in the main text while also providing more detailed text in the unlimited space provided by the online supplement.
guidelines include specific guidance for reporting descriptive statistics, rates, risks, ratios, correlations, regression models, time-to-event analyses, and Bayesian analyses, among many others. If a study includes adjustment for confounding, we prefer that authors show both unadjusted and adjusted results. We also recommend that authors report results on both the absolute (e.g., risk or mean difference) and relative scales (e.g., risk ratio), when relevant. Predicted values (e.g., adjusted probabilities or marginal means) can also greatly aid the interpretation of an analysis and are easily derived in all popular statistical packages (13).

Address the Study’s Limitations Head On
There is not a study design, data source, or statistical method that does not have limitations. We are always impressed when authors present a transparent, balanced, and well-articulated reporting of methods, results, limitations, and potential biases. Conversely, papers that omit, trivialize, or downplay major limitations give a negative impression, and authors will likely be asked to defend these positions empirically.

Acknowledge That Even In-Vogue Methods Have Assumptions and Limitations
We are always excited to review papers that include a thoughtful rationale for the chosen analytic methods (particularly if it educates the readership) and that tests the assumptions of those methods. Sensitivity analyses that demonstrate the robustness of the results by contrasting them with results from alternative analytic strategies or alternative outcome definitions or inclusion criteria are also encouraged.

In our experience, we have seen an increase in manuscripts that use propensity scores and machine-learning methods without the aforementioned qualities. For this reason, we recommend that authors who elect to use propensity scores address at least five specific items; to assist authors, we have added informative citations to the first three recommendations:

1. Explain why propensity score methods are preferable to traditional covariate adjustment (14–17).
2. Explain why a specific propensity score approach (e.g., matching or inverse probability weighting) was used, because the effect estimates have varying interpretations (16–18).
3. Assess the positivity assumption (19).
4. Address limitations intrinsic to all observational studies (e.g., measurement error, confounding by indication, unmeasured confounding).
5. Avoid language suggesting that by performing a propensity score analysis, the observational results are akin to results from a randomized trial.

When authors use machine-learning methods, we expect the same standards in reporting and the use of reporting guidelines (Table 1). For example, the TRIPOD guidelines should be followed if machine-learning methods are used to develop a prediction model. We also challenge authors to justify their methodological choices on the basis of their study question. We are especially wary of manuscripts in which newer methods are used without a clear rationale, those that lack a discussion of the assumptions and limitations of the selected method(s), and those that cannot describe the impact of changes in covariates (i.e., black-box algorithms).

Final Thoughts
Our goal in writing this editorial was to help authors strengthen the presentation of methods and results in research manuscripts. If we had to select only one of these suggestions, it would be that authors find the appropriate guidelines for reporting a particular type of study (Table 1) and ensure that their manuscript includes all of the requested information. The goals of those who create reporting guidelines and checklists are ones that we share: We want manuscripts to describe the different steps of the analysis with sufficient precision to support the 1) reproducibility of the analysis if one had access to the same data and 2) evaluation of the analytic methods and decisions. Relative to the effort required to design a study, collect data, and analyze results, we believe that the limited time required to optimize transparent and thorough reporting of a study is minor, promotes good science, and is likely to create a favorable impression on those reviewing your paper.

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank several colleagues who provided feedback during the development of this editorial, though these individuals bear no responsibility for any errors or omissions in the final draft. Specifically, the authors thank Drs. Andrew D. Althouse, Deepshikha Charan Ashana, Jonathan D. Casey, Gary S. Collins, Scott D. Halpern, Rachel Kohn, Fan Li, Raphael Porcher, and Nadir Yehya. M.O.H. also acknowledges Dr. Meera Nair Harhay and his statistical mentor, Dr. Sarah J. Ratcliffe, for their support in developing the manuscript.

Michael O. Harhay, Ph.D.
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania

and

Statistical Editor, AnnalsATS

Gavin C. Donaldson, Ph.D.*
Imperial College London
London, United Kingdom

and

Statistical Editor, AJRCCM

ORCID IDs: 0000-0002-0553-674X (M.O.H.); 0000-0002-5538-4190 (G.C.D.).

*G.C.D. is Associate Editor of AJRCCM. His participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

References
1. Speckman RA, Friedly JL. Asking structured, answerable clinical questions using the population, intervention/comparator, outcome (PICO) framework. PM R 2019;11:548–553.
2. Davies KS. Formulating the evidence based practice question: a review of the frameworks. Evid Based Libr Inf Pract 2011;6:75–80.
3. Viglianti EM, Admon AJ, Carlton EF, Hensley MK, Prescott HC, Iwashyna TJ, et al. Publishing a clinical research manuscript: guidance for early-career researchers with a focus on pulmonary and critical care medicine. Chest 2019;156:1054–1061.

4. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the “Statistical Analyses and Methods in the Published Literature” or the “SAMPL guidelines.” In: Smart P, Maisonneuve H, Polderman A, editors. Science editors’ handbook. 2nd ed. West Clandon, UK: European Association of Science Editors; 2013 [accessed 2020 Apr 1]. Available from: https://www.equator-network.org/2013/02/11/sampl-guidelines-for-statistical-reporting/.

5. Assel M, Sjoberg D, Elders A, Wang X, Huo D, Botchway A, et al. Guidelines for reporting of statistics for clinical research in urology. BJU Int 2019;123:401–410.

6. American College of Physicians. Information for authors – general statistical guidance. Ann Intern Med [accessed 2020 Feb 27]. Available from: https://annals.org/aim/pages/author-information-statistics-only.

7. Kasza J, Wolfe R, Schuster T. Assessing the impact of unmeasured confounding for binary outcomes using confounding functions. Int J Epidemiol 2017;46:1303–1311.

8. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167:268–274.

9. Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, et al. Control of confounding and reporting of results in causal inference studies: guidance for authors from editors of respiratory, sleep, and critical care journals. Ann Am Thorac Soc 2019;16:22–28.

10. van Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–577. [Published erratum appears in Ann Intern Med 148:168.]

11. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, et al. Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. Crit Care Med 2020;48:623–633.

12. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.

13. Norton EC, Dowd BE, Maciejewski ML. Marginal effects—quantifying the effect of changes in risk factors in logistic regression models. JAMA 2019;321:1304–1305.

14. Leisman DE. Ten pearls and pitfalls of propensity scores in critical care research: a guide for clinicians and researchers. Crit Care Med 2019;47:176–185.

15. Chevret S. Propensity-matching analysis is not straightforward. Am J Respir Crit Care Med 2014;190:362–363.

16. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. J Am Coll Cardiol 2017;69:345–357.

17. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. Stat Methods Med Res 2012;21:273–293.

18. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ 2019;367:i5657.

19. Westreich D, Cole SR. Invited commentary: positivity in practice. Am J Epidemiol 2010;171:674–677. [Discussion, pp. 678–681.]

20. UK EQUATOR Centre. EQUATOR Network: Enhancing the Quality and Transparency of Health Research [accessed 2020 Feb 27]. Available from: https://www.equator-network.org/.