The Value of Statistical Analysis Plans in Observational Research

Defining High-Quality Research From the Start

Laine Thomas, PhD
Eric D. Peterson, MD, MPH

THE INCREASING AVAILABILITY OF ELECTRONIC HEALTH data combined with federal investment has stimulated an expansion in observational clinical research.[1] Observational studies can complement clinical trials and provide important information about comparative safety and effectiveness in populations not well studied in clinical trials. However, there are numerous examples in which the findings from observational studies have failed to be replicated.[2] These failures may be due to several factors, including the exploratory nature of observational questions, failure to fully account for treatment selection bias, known publication biases, and the tendency to pursue post hoc hypotheses. This later problem, termed *data dredging*, is facilitated by the lack of fidelity to a prespecified hypothesis and inadequate reporting of the actual analytic process.

In contrast to observational research, clinical trials ordinarily operate under strict standards at every step of study planning and data analysis. A detailed protocol, including the definition of end points, hypotheses, and all analytical procedures, is submitted to the US Food and Drug Administration and registered in various data repositories, such as clinicaltrials.gov, prior to enrollment of patients. Trial registration helps ensure that both positive and negative findings are publicly known. Prespecification of trial protocols creates an incentive to understand the biological function of the intervention, carefully define the population of interest, target the most appropriate end points, and achieve certainty about the statistical approach. Prespecification of hypotheses and minimal testing means that standard errors and *P* values are accurate measures of uncertainty and statistical evidence is rigorous. Trial protocols can also be referred to and reviewed to understand the questions, end points, and subgroup analyses that were defined ahead of time and those that were post hoc and in need of replication for validation.

See also p 771.

A natural question arises as to whether elements of this rigorous process should be applied to observational research. While select observational studies are already registered in clinicaltrials.gov,[3] some argue that observational research is, by its nature, exploratory and requires substantial flexibility to investigate novel findings and unexpected signals in the data.[4,5] Yet interpretation of statistical evidence (*P* values and confidence intervals) can be made potentially meaningless when multiple hypotheses are generated by exploring the available data. Hence, a balance must be achieved to promote some flexibility but also encourage a rigorous, efficient analytical process that minimizes unnecessary data dredging.

Aside from considering the advantages of preregistration, substantial progress has been made to define standards for reporting observational research.[6,7] The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations provide a checklist of items to include in reporting of cohort, case-control, and cross-sectional studies.[6] Good Research for Comparative Effectiveness (GRACE) principles similarly reflect a consensus on good practice for design and evaluation in comparative effectiveness research.[7] Despite these standards for high-quality observational research and reporting, such guidelines are not consistently adopted, in part because of their complexity and the difficulty of including all components in published articles.

The concepts for improving observational research can be operationalized via use of a formal, prospectively defined statistical analysis plan (SAP). The SAP should include enough detail that another statistician familiar with the data set (or their own independent data) could replicate the analysis. This implies that the SAP should delineate populations (exclusion criteria); end points; descriptive objectives; testable hypotheses; modifications or derivations of standard variables; statistical methods, including handling of missing data, correlated data, bias, and confounding; subgroups; interactions; and sensitivity analy-

Author Affiliations: Department of Biostatistics and Bioinformatics, Duke University School of Medicine (Dr Thomas), and Duke Clinical Research Institute (Dr Peterson), Durham, North Carolina. Dr Peterson is also Contributing Editor, *JAMA*.
Corresponding Author: Eric D. Peterson, MD, MPH, Duke Clinical Research Institute, 2400 Pratt St, Durham, NC 27705 (eric.peterson@dm.duke.edu).

sis. While the SAP should be finalized prior to data analysis, authors may make changes to the analytic plan in response to subsequent findings. These changes to methods, hypotheses, or both should be noted in the SAP to capture when and why components were modified during analysis. Once created, the SAP should be cited in the methods, submitted along with the manuscript for review, and potentially made available as an online appendix to a published article.

The benefits of this process are numerous. First, it promotes good planning rather than haphazard data analysis and communicates this distinction to reviewers and readers. Second, it optimizes statistical resources to focus the best methods on good questions, those with the potential for important findings, either negative or positive. When key hypothesis are defined at the outset, they can be carefully addressed. Third, it facilitates transparency. The existing STROBE recommendations for reporting are quite comprehensive but almost impossible to address in the limited space that is afforded to a published methods section, particularly if the statistical methods are to be replicable. The submission of an SAP provides reviewers with a complete description of what was done, not limited by space. Fourth, this approach may increase efficiency by avoiding distracting messages, maintaining focus on the a priori hypotheses with room for post hoc and sensitivity analysis to be reported.

This suggested process for the SAP in observational research is feasible for real-world adoption. For instance, the Duke Clinical Research Institute, an analytical center for the American College of Cardiology National Cardiac Data Registries, has implemented such an SAP process within their ACTION and Cath/PCI registries. A formal proposal is submitted by clinical researchers, including detailed background and hypotheses, and prioritized by a publications committee. From the proposal, a primary statistician works with the clinical researcher to develop the SAP, translating the clinical questions into descriptive objectives and testable hypotheses. Statistical methods are proposed to address each major objective, including the details mentioned above, table shells for intended output, and the corresponding potential conclusions or takeaway message. A senior statistician and coauthors review the SAP, and it is revised iteratively until all parties support the aims and approach. The review of table shells and cross-checking of potential conclusions with technically stated hypotheses help avoid misunderstandings between the clinician and statistician. These details allow the authors to visualize the project and anticipate issues. While developing the SAP, the statistician may investigate the data availability, extent of missingness, colinearity, and other issues. However, the analysis begins once the SAP is finalized and a single report containing all information is provided to the primary author. Some revisions are nearly always necessary; unforeseen issues with the data may indicate alternative statistical methods or unexpected results may require new analysis. Both the SAP and report are revised to reflect changes.

The process of writing and submitting an SAP captures many of the attributes of clinical trials, without excessive rigidity that would inhibit exploratory research. It requires extra work on the front end but greater efficiency and clarity in producing and reporting results. Current practice may be augmented by making the SAP publicly available as online ancillary material.[8] This would allow readers to confirm and, if desired, replicate the methods used in the study. Some authors have expressed concerns that readers of observational research may become too rigid and dismiss an important finding just because it was not prespecified.[5] However, this can be mitigated if authors make a strong biological case to support post hoc findings and readers may, appropriately, require more confirmatory evidence. The gains in public and academic trust associated with transparency outweigh this concern.[9] Thus, investigators conducting observational research should develop and use prespecified SAPs and should submit these to journals, along with their manuscripts, for review and ultimate online publication.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Patient-Centered Outcomes Research Institute. http://www.pcori.org/what-we-do/. Accessed July 27, 2012.
2. Lauer MS. Will academia embrace comparative effectiveness research? *Acad Med*. 2011;86(6):671-673.
3. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ*. 2010;182(15):1638-1642.
4. Vandenbroucke JP. Observational research, randomised trials, and 2 views of medical science. *PLoS Med*. 2008;5(3):e67.
5. Pearce N. Registration of protocols for observational research is unnecessary and would do more harm than good. *Occup Environ Med*. 2011;68(2):86-88.
6. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806-808.
7. Dreyer NA, Schneeweiss S, McNeil BJ, et al; GRACE Initiative. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care*. 2010;16(6):467-471.
8. Sox HC, Helfand M, Grimshaw J, et al; PLoS Medicine Editors. Comparative effectiveness research: challenges for medical journals. *PLoS Med*. 2010;7(4):e1000269.
9. Ioannidis JPA. The importance of potential studies that have not existed and registration of observational data sets. *JAMA*. 2012;308(6):575-576.