Reflections From a Statistical Editor: Elements of Great Manuscripts

Miguel Marino, PhD

Department of Family Medicine, Oregon Health & Science University, Portland, Oregon
Oregon Health & Science University-Portland State University School of Public Health, Portland, Oregon

I have heard that when bank tellers were originally taught to identify fraudulent bills, they weren't given examples of every possible forgery permutation; instead, they received an authentic dollar and were tasked with diligently studying its hallmarks. Not all genuine dollar bills are alike—some are faded, some wrinkled, some covered with ink stains—but all have an underlying backbone of authenticity that can be spotted after long study. I often think about this when asked to evaluate the statistical rigor of a manuscript being considered for publication.

The *Annals of Family Medicine*, like many other medical journals, doesn't have set rules for the statistical methods a primary care researcher should use and how they should report quantitative analyses. Not having strict rules allows for research creativity, and opens the door for variability in analytic approaches and result reporting, a quality I appreciate as a biostatistician (ie, a devotee of studying variability). Like real dollar bills, there are many unique manuscripts that are authentically great; like counterfeit ones, there are also those that miss the mark.

As with a dollar bill, there are many little details that make up an authentically good manuscript: an informative title, an abstract that captures the entire story, self-sufficient and well-labeled figures and tables, etc. While some qualities of great publications are well-known, I'd like to highlight some from a statistical perspective that I have encountered as *Annals'* statistical editor.

Conflicts of interest: Author is the Statistical Editor for Annals of Family Medicine*.*

CORRESPONDING AUTHOR

Miguel Marino, PhD
Department of Family Medicine
Oregon Health & Science University
3181 SW Sam Jackson Park Rd
Portland, OR 97239
marinom@ohsu.edu

ADDRESS DIFFERENTIAL DROPOUT

In randomized trials or observational studies comparing 2 or more groups, there are almost always patients, clinics, etc who drop out of the study, for a variety of reasons. Dropout in longitudinal studies is a potential source of bias and should be addressed in any manuscript. Despite CONSORT statements and extensive literature describing how to address dropout, there are still 2 common misconceptions: (1) If dropout is similar between groups, bias is not a concern; and (2) If dropout is different between groups, bias exists. Bell et al[1] investigated these misconceptions and concluded that any dropout bias depends on the intervention effect being estimated, the mechanism resulting in the missing data (ie, missing completely at random (MCAR), missing at random (MAR), not missing at random (NMAR)), and the study's analytic approach.

Under certain scenarios, differential dropout can still yield unbiased results. For example, if data are MCAR or MAR and a likelihood-based mixed model is utilized, differential dropout may produce unbiased estimates because the model borrows information from completely observed subjects to implicitly impute missing values. Conversely, Bell et al showed that equal dropout between groups can still produce biased intervention effects. The strongest manuscripts provide detailed information about dropout over the course of the study, evaluate missing data mechanisms, select appropriate analytic models that can flexibly handle dropout, and perform sensitivity analyses when dropout is anticipated to be NMAR.

STRIKE A BALANCE BETWEEN DETAILED AND SUCCINCT REPORTING

Pragmatic studies are foundational in primary care research. They allow researchers to investigate how effective an intervention or treatment strategy is in the context of everyday practice. When conducting

a pragmatic study, numerous decisions are made that may affect its internal and external validity. Regardless of these decisions, great manuscripts find the balance between reporting enough detail for readers to make informed judgments about the study validity, while keeping the manuscript succinct. It has been shown that poorly reported studies are more likely to be biased than well-reported research.[2]

In pragmatic studies, protocol deviations (ie, deviations from what was originally planned) can occur to meet the realities of everyday practice. When protocol deviations differ between 2 or more groups being compared, there is a potential for bias.[3] Researchers should investigate and track protocol deviations and account for them when possible. Results from intent-to-treat analyses should also be reported with per-protocol results to evaluate the extent of bias from excluding non-protocol–adherent patients or practices. Great manuscripts result from studies with protocols that closely fit normal practice and with participants that try to adhere to the protocols; these papers also provide details about observed protocol deviations and investigate reasons for non-adherence and potential bias, all while being succinct. If more detail is needed than the sophisticated but non-expert reader might want, additional specifics can be included in an appendix.

ACCOUNT FOR CLUSTERED DATA

One defining feature of primary care research is the setting of our studies. Many studies evaluate outcomes and interventions that are relevant to patients continually seeing clinician(s) working in a practice or network of practices. The study settings produce a natural set of clusters (multiple visits nested within patients, nested within clinicians, nested within states/regions) that challenge the traditional assumptions and machinery of classical statistical methods like linear and logistic regression.[1] Ignoring the clustered nature of this type of data can result in a lower effective sample size, resulting in lower power to detect meaningful effects.

In great manuscripts, the multilevel nature of this data clustering is appropriately accounted for (eg, mixed effects modeling, Generalized Estimating Equation modeling) and a measure of the unit-relatedness within clusters, such as the intracluster correlation coefficient (ICC), are reported.[4] Reporting the ICC and other similar measures allows primary care researchers a glimpse into whether an outcome is more of a patient- or practice-level construct, and allows them to use these measures for future study planning.[1]

TAKE ADVANTAGE OF THE STUDY DESIGN IN THE ANALYTIC APPROACH

Recruiting a large number of practices for a primary care research study is difficult.[5] As such, in cluster randomized trials, it is rare to see studies with a large enough sample of practices for simple randomization to yield appropriate balance in observed and unobserved confounders between study groups—simple randomization mostly works with large samples. Many primary care researchers smartly circumvent this limitation by performing stratified or block randomization, schemes that force confounding factors to be balanced between study groups. When this is done, literature suggests that ignoring variables that were used in stratified or block randomization procedures in the analytic phase of the study results in CIs that are too wide and *P* values that are too large.[6] The gains in statistical power achieved by balancing important confounders in the study design phase are lost when analytical methods do not account for strata/block variables, and a treatment or intervention that may be beneficial could be incorrectly classified as ineffective.

In observational studies, there are similar parallels about the importance of tailoring the analytic approach to incorporate decisions made in the study design phase. For example, in observational studies that utilize propensity score matching to balance observed confounders between study groups, some methodologists recommend using analytic approaches that account for the paired nature of the data resulting from a propensity score matching procedure.[7] Great manuscripts utilize analytic approaches that match the study designs, maximizing the reason and benefits for such designs.

DON'T OVERESTIMATE PREDICTION PERFORMANCE

With the advent of "big data," primary care researchers started to collect more data in both dimensions (number of patients/practices, and number of patient/practice characteristics), which has resulted in an explosion of prediction models which try to take advantage of these "big data" to assess future events that could affect current clinical decisions. Predictive models that use a set of data to build the model, however, and then use the *same* set of data to assess how well their model performs will overestimate the prediction accuracy.[8]

Great manuscripts use prediction models that include important predictors, assess prediction performance in their own data, but more importantly also assess external validation performance (either on a completely new set of patients/practices or with an internal cross-validation procedure) to ensure that the

models are not overestimating. Lastly, for prediction models to be useful in primary care, researchers should start addressing how a busy clinic could implement these prediction models and how they may use this information in their everyday practice.

TREAT OBSERVATIONAL STUDIES LIKE RANDOMIZED TRIALS

In evidence-based medicine, the randomized trial is still the gold standard for determining causal inference. If 3 new treatment options are available, a randomized trial comparing all possible treatments (and potentially a control treatment) would help address comparative effectiveness and determine which treatment works for patients and under what settings. Unfortunately, randomized trials are expensive and cannot possibly be done for all possible treatments or interventions. When randomized trials are not possible, a rigorous observational study is the next best option.

In general, there is always concern that observational studies are prone to many types of bias. Recent work[9,10] has shown that when researchers design observational studies to emulate a hypothetical randomized trial (ie, mimicking the design features and protocols of a true randomized experiment), this can lead to unbiased treatment effects. For strong observational studies and manuscripts when randomized experiments are not possible, conceptualize what kind of randomized trial would have been performed (eg, identify eligibility criteria, follow-up time, outcomes of interest, causal effects of interests such as intent-to-treat or per-protocol, analysis plan, etc) and then implement this plan using their observational study with the goal of producing results that are similar to observing this hypothetical randomized trial. Research reporting guidelines,[2] such as the STROBE reporting guidelines, provide helpful guidance.

Primary care research is continually evolving, and the components that make up a great manuscript are similarly changing to match the demands of our field. While the list here is not exhaustive, it represents some core topics that may help the reader turn a promis-

ing project into a great manuscript. It is my hope that primary care researchers will continually improve the rigor and peer-review of our quantitative research and learn to distinguish between the features of authentic and "counterfeit" manuscripts.

To read or post commentaries in response to this article, see it online at http://www.AnnFamMed.org/content/15/6/504.

References

1. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential drop-out and bias in randomised controlled trials: when it matters and when it may not. *BMJ.* 2013;346:e8668.

2. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352(9128):609-613.

3. Fransen GA, van Marrewijk CJ, Mujakovic S, et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. *BMC Med Res Methodol.* 2007;7:16.

4. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Ann Fam Med.* 2004;2(3):204-208.

5. Sweeney SM, Hall JD, Ono SS, et al. Recruiting practices for change initiatives is hard: findings from EvidenceNOW [published online ahead of print Sep 4, 2017]. *Am J Med Qual.*

6. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ.* 2012;345:e5840.

7. Haukoos JS, Lewis RJ. The propensity score. *JAMA.* 2015;314(15): 1637-1638.

8. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning.* New York, NY: Springer; 2013.

9. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766-779.

10. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol.* 2017;32(6):495-500.

11. Zyzanski SJ, Flocke SA, Dickinson LM. On the nature and analysis of clustered data. *Ann Fam Med.* 2004;2(3):199-200.

12. Orkin AM, Phillips WR, Stange KC. Research reporting guidelines and the new Annals instructions for authors. *Ann Fam Med.* 2016;14(6):500-501.