What Do ICU Clinicians Really Need to Know About Statistics: Time to Give Up or Time to Bridge the Gap Between the Evidence and the Reader?*

I must confess to always having viewed studying statistics as similar to a screening colonoscopy; I knew that it was important and good for me, but there was little that was pleasant or fun about it.–Anthony N. DeMaria, MD, MACC

Daniel E. Leisman, BS
Icahn School of Medicine at Mount Sinai
New York, NY
Feinstein Institute for Medical Research
Manhasset, NY

The above was how Dr. DeMaria, then Editor-in-Chief of the prestigious *Journal of the American College Cardiology*, once began a talk. The joke was a hit and saliently relatable for the clinical audience. Yet, as much as clinicians love to hate statistics, understanding statistics turns out to be one of those things that can matter at the bedside.

The singular purpose of medical research is to improve the lives of patients. We conduct, report, debate, and implement research to provide patients the care most likely to achieve the best result. Statistics is the language of medical research. Clinical experiments are designed, analyzed, and reported within statistical frameworks. Doing right by patients involves applying evidence to their unique contexts of chronic health, acute physiology, and values. Understanding what evidence means, and how it applies (or not) for a given patient, invariably requires some facility with statistics.

Yet, trainees and senior clinicians alike frequently struggle with core-concepts. Consider how often presenters, whether at morning conference or international meetings, hand-wave through statistical procedures. Worse, erroneous designs or interpretations can go unchallenged when clinicians and statisticians feel uncomfortable in each other's domains. Rampant inappropriate methodology and improper interpretation in medical literature have been recognized for decades (1).

We see a prominent example of this disconnect in the widespread misunderstanding and misuse of *p* values (2). The myriad misconceptions clinicians commonly hold about *p* values are extensively discussed (3). *p* values are not intrinsically flawed when used properly, but they become deeply problematic when misinterpreted, such as when arbitrary *p* value of less than 0.05 thresholds entirely dictate conclusions. Due to this widespread practice and many others (3), experts have long warned against using *p* values as we often encounter them (1). These recommendations subsequently formalized as explicit guidance for randomized trials by the Consolidated Standards of Reporting Trials group in 2010, and a comprehensive statement on *p* values in 2016 from the American Statistical Association (ASA) (4, 5). Nevertheless, use, and misuse, of *p* values continued to increase from 1995 to 2015 (6). Although we do not yet know if *p* value reporting has improved since the ASA statement, anecdotally, the problem likely persists (7, 8). Lack of consensus on alternatives exacerbates challenges in eschewing *p* values. Proposals include reporting only effect-sizes and uncertainty estimates, lowering *p* value thresholds, Bayesian inference, and "second-generation" *p* values (5, 9, 10). Each carries problematic implications, although each is also almost certainly much better than the current situation.

On this foundation, and fueled by publish-or-perish incentives, emerges what the late eminent biostatistician, Doug Altman, PhD (1), called, "The Scandal of Poor Medical Research." Pervasive misconceptions and even occasional impropriety produces a landscape of inappropriate use, reporting, and interpretation of statistics in biomedical research, with consequences increasingly coming to light. "The Reproducibility Problem" is a euphemism for the fact that many observations in our evidence-base cannot be replicated. For example, in 2012, an industry-funded project attempting to replicate 53 landmark preclinical cancer studies made headlines when it could reproduce just six (11%) (11). We have seen countless clinical studies where effects diminished or even reversed in larger-scale, more rigorous investigations. Clinicians should care deeply about this because it impacts patients. A misstep in patient care may harm someone. A blunder in research or interpretation may harm thousands.

For clinicians, the singular question amid this controversy remains simple: what do I need to know to give the patient in front of me the best possible care?

Enter the study by McCullough et al (12), published in this issue of *Critical Care Medicine*. The authors take an important first step, asking what methods are encountered in recent high-impact critical care randomized controlled trials (RCTs): that is, what would critical care clinicians need to know about statistics to interpret RCTs? To answer this, the authors conducted a systematic review of RCTs in the MEDLINE database

***See also p. 1985.**

Key Words: biostatistics methods; clinical trials, randomized; data interpretation, statistical; statistical data analysis; medical education; model, statistical

Mr. Leisman has disclosed that he does not have any potential conflicts of interest.

DOI: 10.1097/CCM.0000000000003420

published in ten high-impact journals from 2011 to 2015. They identified 116 original articles relevant to critical care and recorded reported statistical procedures with exceptional granularity. They also compared their observations to earlier literature, showing how these methods' use has evolved over time.

In short, the authors found frequent and rising use of advanced statistical methods in critical care RCTs. They identify several increasingly used domains. Some readers may readily understand the most common and simple: contingency, *t* tests, and analysis of variances. Other procedures are fairly straightforward but pervasively misunderstood despite common use: Kaplan-Meier statistics and multivariable linear, logistic, and proportional-hazard regressions (13). This is important because expanding clinicians' proficiency with these common methods and their underlying assumptions could prove low-hanging fruit for intervention. However, some increasingly prevalent techniques sit firmly in the realm of advanced methods. Half the RCTs included generalized linear or mixed-effects models for longitudinal/correlated data. These underlying mechanics are likely beyond what one could reasonably ask nonstatisticians to learn, raising questions of how to include conceptual understanding of these procedures in the clinician's toolbox.

The authors also detect a troubling pattern. Forty-percent of RCTs reported *p* values for descriptive characteristics. This is consistent with analyses of wider biomedical literature (6) and confirms critical care is not insulated from *p* value misuse. Readers should recognize *p* value "columns" are misleading, look instead for effect-sizes, and realize high-impact journals and prestigious affiliations do not guarantee proper methodology.

Notably, these findings are almost certainly understated. First, they only analyzed RCTs. Of course, RCTs are not the only way practice-changing data are generated. Many influential studies for day-to-day practice answer observational questions where randomization is inappropriate. Since RCTs are often more quantitatively straightforward than observational studies, many important methods are likely underrepresented or not captured if attempting to generalize this analysis to all critical care literature. Second, the authors may have selected for higher quality papers. RCTs published in outlets other than the 10 high-impact journals could be even more likely to improperly report *p* values. Third, the authors did not identify any trials with Bayesian adaptive designs, which have been used for years in fields like oncology to improve the efficiency and ethics of clinical trials. These designs are particularly relevant to critical care, and several such trials are currently in-progress (14, 15). Critical care clinicians should soon expect to encounter this additional complexity in their literature.

Nonetheless, the authors clearly show critical care trials use a lot of complex math. Command of the evidence-base is perhaps more imperative in the ICU, where stakes are higher for patients and families. How should we confront this barrier to providing evidence-based care for practitioners without statistical education?

The authors offer their idea for a solution. They conclude, "In addition to [training] clinicians in relevant methodology... specialist biostatistical support [is] integral to… clinical research [and] evidence-based clinical practice." They note that despite improving clinician familiarity with biostatistics, increasing methodological sophistication likely outpaces these improvements, creating accessibility barriers at levels as basic as vocabulary. The authors' implication is that things are simply too complicated for clinicians to understand, and bedside biostatistical support might more feasibly bridge this gap. They may be right.

However, although the author's proposal has merits, we should consider some drawbacks to statistical consultation for clinical care. First, it is unknown if this approach would actually make care more evidence-based, let alone enough to improve outcomes. Second, it is hard to imagine how this could be feasibly implemented. Would biostatisticians come to morning rounds? Perhaps their expertise would be sought as a consult. Would clinicians use this service? Will clinicians know when they need assistance interpreting evidence? Data suggest otherwise (13). Paraphrasing Mark Twain, "What gets people into trouble ain't what they don't know, but what they know that just ain't so." Third, we should recognize Ivory Tower solutions. How many biostatisticians are available to work in ICUs of small rural communities or low and middle-income nations?

But, most importantly, we should ask if we are ready to give up on achieving adequate biostatistical competence among clinicians. Decision-makers should use others' expertise, but should also be able to interpret the data influencing their decisions. Is there really that much difference between being equipped to interpret the literature versus an electrocardiogram or ultrasound? These skills are supplemented by cardiology and radiology consults, but they are worth acquiring because we think they improve patient care. Certainly, understanding evidence falls in that category too.

What should we take away from this article (12)? First, we see prevalent *p* value misuse in critical care trials. Editorial boards should take firm stances to correct this, especially if readers cannot reliably interpret *p* values correctly. Second, this report suggests widening gaps between clinicians' statistical literacy and the methods of the evidence-base. Failure to bridge these gaps will render clinical reports inaccessible to their intended audience. We must redouble efforts to help students, trainees, and lifelong learners achieve biostatistical competence and support. What specific content constitutes adequate statistical competency for critical care, what interventions improve competency, and what impact competency has on evidence uptake are questions worth asking. McCullough et al (12) provide a reasonable starting point for the former.

Ultimately, these findings get us a bit closer to knowing what we need to know and remind us that even when neither fun nor pleasant, statistics are important and good for us.

REFERENCES

1. Altman DG: The scandal of poor medical research. *BMJ* 1994; 308:283–284
2. Kyriacou DN: The enduring evolution of the p value. *JAMA* 2016; 315:1113–1115
3. Goodman S: A dirty dozen: Twelve p-value misconceptions. *Semin Hematol* 2008; 45:135–140
4. Schulz KF, Altman DG, Moher D; CONSORT Group: CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340:c332
5. Wasserstein RL, Lazar NA: The ASA's statement on p-values: Context, process, and purpose. *Am Stat* 2016; 70:129–133
6. Chavalarias D, Wallach JD, Li AH, et al: Evolution of reporting p values in the biomedical literature, 1990–2015. *JAMA* 2016; 315:1141–1148
7. Gates S: New England Journal of Medicine for the last time. Reducing Uncertainty: clinical trials and statistics, mostly. 2018. Available at: https://simongates.com/2018/08/01/new-england-journal-of-medicine-for-the-last-time/. Accessed September 5, 2018
8. The New England Journal of Medicine: New Manuscripts: Guidelines for Statistical Methods. 2018. Available at: https://www.nejm.org/author-center/new-manuscripts. Accessed August 4, 2018
9. Ioannidis JPA: The proposal to lower p value thresholds to.005. *JAMA* 2018; 319:1429–1430
10. Blume JD, D'Agostino McGowan L, Dupont WD, et al: Second-generation p-values: Improved rigor, reproducibility, & transparency in statistical analyses. *PLoS One* 2018; 13:e0188299
11. Begley CG, Ellis LM: Drug development: Raise standards for preclinical cancer research. *Nature* 2012; 483:531–533
12. McCullough JPA, Lipman J, Presneill JJ: The Statistical Curriculum Within Randomized Controlled Trials in Critical Illness. *Crit Care Med* 2018; 46:1985–1990
13. Windish DM, Huot SJ, Green ML: Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* 2007; 298:1010–1022
14. Lewis RJ, Angus DC, Laterre PF, et al: Rationale and design of an adaptive phase 2b/3 clinical trial of selepressin for adults in septic shock. Selepressin evaluation programme for sepsis-induced shock-adaptive clinical trial. *Ann Am Thorac Soc* 2018; 15:250–257
15. REMAP-CAP: Randomized, Embedded, Multifactorial Adaptive Platform Trial for Community-Acquired Pneumonia, 2016. Available at: https://clinicaltrials.gov/ct2/show/NCT02735707. Accessed August 4, 2018

Central Venous Catheter Failures: Nowhere Near Zero*

Naomi P. O'Grady, MD
Internal Medicine Service
Clinical Center
National Institutes of Health
Bethesda, MD

Sameer S. Kadri, MD, MS
Critical Care Medicine Department
Clinical Center
National Institutes of Health
Bethesda, MD

Central venous access devices (CVADs) offer a critical conduit for the delivery of life-sustaining therapies to patients in the ICU. The insertion and maintenance of CVADs are associated with considerable morbidity, including, but not limited to, central line-associated bloodstream infection (CLABSI), local infection, thrombosis, and dislodgement. Evidence, or even suspicion of these complications, often results in early removal and replacement of the CVAD, exposing the patient to additional procedural risks. Much attention has been given in recent years to quantifying and preventing CLABSIs (1, 2). However, in the literature, there is a conspicuous paucity of estimates for CVAD failures (apart from CLABSIs) across all types of catheters. These data are essential to gauge patients' overall burden of risk from such devices.

In this issue of *Critical Care Medicine*, Takashima et al (3) evaluated CVAD failures and complications across the gamut of CVADs in the ICU population by way of meta-analysis spanning 63 studies containing 50,000 CVADs and 396,951 catheter days. The review included randomized controlled trials and observational cohort studies dating back to 2006 and enrolling adults with CVADs in the ICU setting that reported the outcomes of interest. The primary outcome for this meta-analysis was CVAD failure, defined as removal of the device prior to the completion of therapy. The secondary outcomes were CVAD complications after successful insertion defined as CLABSI, catheter-related bloodstream infection, catheter-associated venous thrombosis, removal due to suspected infection, occlusion, dislodgement, breakage, local infection, or phlebitis. The authors followed standard operating procedure for systematic reviews which included careful study selection, a priori reporting of study design and rigorous assessments of study quality, risk of bias, and heterogeneity where possible.

The study by Takashima et al (3) identified that one in every 20 CVADs fail (5%) before the completion of treatment in the ICU. This is an important finding not only in the context of patient safety, but also from the perspective of provider time, hospital resources, and healthcare dollars associated with

***See also p. 1998.**

Key Words: central line-associated bloodstream infection; central line-associated bloodstream infection rates; central venous catheter; complications; intensive care unit; vascular access

The findings and conclusions in this article are those of the author and do not necessarily represent the views of the National Institutes of Health, or the Department of Health and Human Services.

Supported, in part, by the National Institutes of Health, Clinical Center.

Drs. O'Grady and Kadri received support for article research from the National Institutes of Health and disclosed government work.

DOI: 10.1097/CCM.0000000000003479