

Replication to Advance Science

Changes in ANESTHESIOLOGY

James C. Eisenach, M.D., Timothy T. Houle, Ph.D.

“WHAT’S new with you?”
“Take a look at that!” “Well,
that’s certainly different!”

We are interested in novelty and are wired to notice changes in our environment. Advances in technology mean that many of us are more frequently prodded with new stimuli with each passing day. In this issue of *ANESTHESIOLOGY*, we highlight something which appears on face value not to be new. A group of investigators replicated a score to predict postoperative pulmonary complications in a different group of patients.¹ “How boring!” you might think. Why is this in a journal that focuses on new discoveries? Let us briefly describe why this article is important and steps this journal is taking to encourage more articles like it.

Judging from the number of this type of publications in the specialty, researchers at least are very interested in predicting postoperative complications. Defining risk factors for postoperative complications is important because these factors can provide clues to guide fundamental science research on mechanisms for these complications. From a clinical perspective, knowing with reasonable precision that a patient is at high risk for major complications allows for a more informed discussion with the patient so that they can appreciate and weigh risks and benefits to the surgical procedure. Certain risk factors are amenable to preoperative treatment or correction, and precautions can be taken in high-risk patients to either start preventive therapy or more intensely monitor for the onset of complications and begin treatment early.

Despite the popularity of risk scoring articles and their logical application to clinical practice, predictive models are rarely validated beyond the study from which they were



“To best understand replicability, consider the various degrees of replication as rungs on a ladder.”

devised and even more rarely meet their full clinical potential. A key problem in the use of predictive scores is their lack of predictive accuracy in future patients and different settings. Even with a study of hundreds or thousands of patients, the apparent validity of a predictive model is optimistic when compared with how it will perform in future applications. This is true because the model has been specified (*i.e.*, which predictors?) and estimated (*i.e.*, how much to weight each predictor?) on that particular study group, and the likelihood of similar high performance, even at the same institution in similar patients, is remote.

Statistical approaches such as split-sample, cross-validation, and bootstrapping are all elegant forms of internal validation procedures that can be used to estimate a model’s predictive value in future samples.² Some form of internal validation procedure is required for publication of novel prediction models in *ANESTHESIOLOGY*. However, these efforts are only estimates and should be viewed as merely a starting place for the validation process. What is really needed is additional work, preferably by outside researchers across different institutions and settings. In short, a prediction study needs to be replicated.

To best understand replicability, consider the various degrees of replication as rungs on a ladder. The first rung of the ladder is reproducibility. The concept of reproducibility refers to the ability to confirm a research finding using the same data set.³ Stated simply, a finding is reproducible if it can be confirmed by a second researcher when given access to the same data as that of previous researcher. Reproducibility is a necessary but not sufficient step for the

Image: ©Thinkstock.

Corresponding article on page 219.

Accepted for publication May 14, 2014. From the Department of Anesthesiology, Wake Forest School of Medicine, Winston-Salem, North Carolina.

Copyright © 2014, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. *Anesthesiology* 2014; 121:209-11

second rung of the ladder, replicability. Replicability refers to obtaining the same finding in multiple random samples that share the same core elements.³ This would be akin to finding the same level of predictive accuracy of a model that used patient-level predictors to predict mortality in two different intensive care units. Replicability is a necessary but not sufficient step for the final rung of the ladder, generalizability. Generalizability refers to obtaining the same finding in multiple random samples when even core elements of the studies differ.³ An extreme example of this would be to evaluate the intensive care unit model in a postanesthesia care unit setting where patients and procedures differ greatly. A model has the potential for high generalizability when it does not depend on unmeasured variables for performance. Examining generalizability requires that researchers seek out a diversity of settings or influences that could affect their findings.

This process is precisely what was done in the study by Mazo *et al.*,¹ who recruited patients from more than 63 centers across 21 countries in Europe to determine whether the predictive model they had created from study of patients in Catalonia in Spain⁴ could be replicated/generalized in these settings. Not surprisingly, the performance of the predictive model was not as good as in the original dataset from which it was optimized, but it still functioned at a moderate to good level in predicting postoperative pulmonary complications. Also not surprisingly, they observed that performance of the model was not uniformly generalizable across all centers—it was much weaker and less useful in hospitals in Eastern Europe, for example, suggesting that other factors not included in the model play an important independent or moderating role.

The importance of this article resides not in the model's specificity and sensitivity—it works only reasonably well most of the time—but in the process of externally validating the model. We publish predictive score models in the hope that their true utility to influence clinical practice will be later defined by replication. Sadly, the current article is one of the very few instances where this hope has been fulfilled.

Replication of research is essential to scientific advancement, but with our focus on novelty, raising enthusiasm among investigators to reproduce and replicate research and among journals to publish replication work is difficult. At a joint meeting of the Editorial Boards of *ANESTHESIOLOGY* and *Anesthesia & Analgesia* (October 14, 2013), John Ioannidis, M.D., D.Sc. (Professor of Health Research and Policy, Stanford University, Stanford, California), a leader in reproducible science, reviewed the classic cycle of new discoveries in medical science. The first study of a new idea or therapy, typically with a small number of research subjects, observes a very large (usually positive) effect and is published in a high-impact journal with great fanfare. Subsequent studies note considerably smaller effects and adverse effects not noted in the initial report, leading to a synthesis of the literature which

concludes that the benefit of the new idea or treatment is small and must be weighed against these adverse events.

Dr. Ioannidis challenged us to speed the process of maturing our understanding of new advances by encouraging the publication of replication work, and the work by Mazo *et al.*¹ is such an example. Although we aim to publish important new ideas in this journal, we have embarked on an education effort within the Editorial Board to recognize the importance of well-performed replication studies such as this one, and our hope is that you will see more such studies published in the journal.

In a broader sense, there is an increasing concern that major findings cannot be reproduced by other investigators, especially in observational clinical studies and preclinical studies. In response to this concern, workgroups in these two areas were established and presented recommendations to the Editorial Board on methods to increase transparency of reporting to thereby facilitate the ability of other investigators to reproduce the published work. Future editorials, accompanied by changes in Instructions to Authors and policies regarding transparent reporting will appear in the coming months regarding articles using these research methods.

Transparent reporting of randomized controlled trials is mature, and we use custom software to scan all articles of this type to assure that they meet the elements of reporting recommendations which we feel are essential for this class of article. Yet, reporting of full details of methods remains incomplete, and we have heretofore had no formal policy regarding sharing data from such studies. As noted above, reproduction and replication are essential to refining our understanding of effect size, and to do this, access to original data is necessary for the first and extremely helpful for the second. Despite calls by federal funding agencies and the public to accessibility of data, there are many barriers to such sharing original data, and journals are not ideal hosts to archive such data.

In an effort to encourage reproduction and replication, we are following the lead of *Annals of Internal Medicine* and will shortly add a section at the end of articles called Reproducible Research. This will include two statements; whether the full protocol and statistical code are available and whether the original data (de-identified to protect patient privacy) are available and if so, the contact information. As is the case at *Annals of Internal Medicine*,⁵ these statements will have no bearing on the peer-review process or the decision to accept articles. Authors will not be forced to provide full access to either the protocol or original data. In addition, the extent and conditions under which these materials will be made available can be specified, such as after all planned primary and secondary analyses are completed, or after a certain date, for example. This policy has been well received by authors and investigators at *Annals of Internal Medicine* (personal communication, Christine Laine, M.D., M.P.H., Editor-in-Chief, March 20, 2014), and we hope

that this step will increase the likelihood that important work in *ANESTHESIOLOGY* will be reproducible and replicated.

In summary, our focus on novelty in science and publishing neglects the essential role that research replication has in advancing ideas and improving clinical care. The replication/validation study by Mazo *et al.*¹ represents an important advance in predicting postoperative pulmonary complications and serves as a good illustration of the value of replication. We encourage submissions of well-performed replication/generalizability studies. We will also start providing contact information on the availability of complete protocols and original data to further encourage reproducible research.

Competing Interests

Dr. Eisenach has received fees for consultation to Aerial Biopharma (Morrisville, North Carolina) and Adynxx (San Francisco, California), and Dr. Houle has received research funding from Merck Sharp & Dohme (Whitehouse Station, New Jersey) and DepoMed (Newark, California) on topics unrelated to this editorial.

Correspondence

Address correspondence to Dr. Eisenach: editor-in-chief@anesthesiology.org

References

1. Mazo V, Sabaté S, Canet J, Gallart L, Gama de Abreu M, Belda J, Langeron O, Hoefft A, Pelosi P: Prospective external validation of a predictive score for postoperative pulmonary complications. *ANESTHESIOLOGY* 121:219–31
2. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD: Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54:774–81
3. Asendorpf JB, Conner M, De Fruyt F, De Houwer J, Denissen JJA, Fiedler K, Fiedler S, Funder DC, Kliegl R, Nosek BA, Perugini M, Roberts BW, Schmitt M, van Aken MAG, Weber H, Wicherts JM: Recommendations for increasing replicability in psychology. *Eur J Pers* 2013; 27:108–19
4. Canet J, Gallart L, Gomar C, Paluzie G, Vallès J, Castillo J, Sabaté S, Mazo V, Briones Z, Sanchis J; ARISCAT Group: Prediction of postoperative pulmonary complications in a population-based surgical cohort. *ANESTHESIOLOGY* 2010; 113:1338–50
5. Laine C, Goodman SN, Griswold ME, Sox HC: Reproducible research: Moving toward research the public can really trust. *Ann Intern Med* 2007; 146:450–3