

Study: Subpar reporting of biomarker characteristics

A contributor to the reproducibility crisis?

Anne Paxton

January 2020—Since the *Journal of Irreproducible Results* arrived on the scene in 1955, founded by a physicist and a virologist, its parodies of scientific research have amused and skewered many in the scientific community.

But in the world of clinical research, irreproducibility is a less than whimsical idea. Over the past 20 years, the increasing number of clinical studies with findings that are not confirmed upon retesting has created a “massive replication crisis” in clinical medicine and preclinical research, writes Patrick M. Bossuyt, PhD, professor of clinical epidemiology at the University of Amsterdam, in a recent *Clinical Chemistry* editorial titled “Laboratory measurement’s contribution to the replication and application crisis in clinical research” (2019;65[12]:1479–1480).



‘If the journals in their instructions to authors insist that this information be provided, then authors will provide it.’ —

David Sacks, MB, ChB

Researchers at drug companies report anecdotally that a relatively small fraction of published findings are reproducible when investigated, says David B. Sacks, MB, ChB, senior investigator and chief of the clinical chemistry service in the Department of Laboratory Medicine at the National Institutes of Health. A few years ago, Dr. Sacks wondered how laboratory issues might factor into the reproducibility problem and decided with colleagues to take a look. That led to a startling discovery.

“Measurement of biomarkers is fundamental to many, many aspects of patient care, ranging from diagnosis of patients to management and evaluation of treatment,” Dr. Sacks says. But “I noticed that many clinical papers didn’t describe how they measured biomarkers really well. And I realized that nobody had ever addressed the lab aspects.” He and colleagues at the NIH, University of Virginia, and Weill Cornell Medicine in New York undertook a study of the laboratory testing aspect of clinical studies involving biomarkers, which often form the basis for clinical guidelines.

When the authors examined 544 studies (and 1,299 biomarker uses) published in the top five clinical medical journals, the level of inadequate reporting took them aback. Their study found that “reporting of the analytical performance of biomarker measurements is variable and often absent from published clinical studies” (Sun Q, et al. *Clin Chem*. 2019;65[12]:1554–1562).

“I was shocked by the magnitude of the omissions we observed,” Dr. Sacks says. For two-thirds of the biomarkers,

no information on analytical characteristics was provided, while for a majority of biomarkers the manufacturers could not be identified and there was no information about the trueness or precision of the methods used to measure the biomarker.

Inadequate reporting is a threat to interpretation and replication of study findings, Dr. Sacks and his coauthors write. As Dr. Bossuyt confirms in his editorial, insufficient, uninformative reporting of research methods “renders all attempts at reproducing the original findings hazardous” and is one issue that returns repeatedly in all explanations of the massive replication crisis.

The Sun, et al., study examined articles published between 2006 and 2016 in *Annals of Internal Medicine*, *JAMA*, *Lancet*, *New England Journal of Medicine*, and *PLOS Medicine*. If a biomarker was used for selection or classification of participants or as a study outcome, or for any combination of those uses, the study was selected. Articles were excluded if they were not clinical studies in humans, had fewer than 10 participants, used only immunohistochemical or imaging markers, were meta-analyses of primary studies or computer modeling studies with no actual biomarker measurements, or had been retracted.

The most frequent molecular biomarker types were proteins (55 percent), nucleic acids (12 percent), and lipids/steroids (12 percent). The most common biomarkers were C-reactive protein (six percent), cardiac troponin (five percent), and glucose (four percent).

Eleven key analytical characteristics were evaluated: accuracy, day-to-day imprecision, within-run imprecision, otherwise unspecified imprecision, analytical sensitivity, interferences, reportable range of results, reference interval, cutoffs for test positivity or decision limits, quality control, and calibration/calibration verification. The authors assigned a score of zero to 11 for each biomarker based on the number of analytical performance characteristics reported. For 865 of the 1,299 biomarker measurements, the score was zero. The median reporting rate for each of the five journals and for the full data set of analytical characteristics was also zero. Of the analytical characteristics evaluated, the reporting rate of total imprecision was the highest, at 13 percent; that of interference studies was the lowest, at two percent.

This low reporting occurred despite the known fact that for many biomarkers, results for a given patient sample differ depending on the manufacturer of the test. For example, the authors point out, cardiac troponin I results have varied as much as 33-fold for some assays. Other manufacturer-dependent differences are seen with thyroid stimulating hormone, prostate-specific antigen, and human chorionic gonadotropin, as well as long established biomarkers such as albumin and creatinine. Without identification of the manufacturer, the study says, clinicians “cannot know if the decision cutpoints used in the study are appropriate for the test in their hospitals or at other laboratories where their patients’ samples are assayed.”

The kinds of problems a lack of reproducibility can create go far beyond the theoretical, as the Sun, et al., study explains in relation to tight glycemic control. The 2001 van den Berghe trial, a randomized controlled study that concluded that tight glycemic control decreased mortality in critically ill patients, led to a massive shift in clinical guidelines for treatment (*N Engl J Med.* 2001;345[19]:1359–1367). But a second trial, reported in 2009 (the multicenter NICE-SUGAR trial), appeared to contradict the results of the van den Berghe trial, finding an increase in mortality with tight glucose control (*N Engl J Med.* 2009;360[13]:1283–1297).

As the Sun, et al., study notes, the van den Berghe trial measured glucose by a highly accurate and precise method, described in the publication, while the NICE-SUGAR study included glucose measured by a variety of methods, including point-of-care glucose meters. No manufacturers’ names were provided in the second study, nor was there information about the analytical performance of the various methods. In addition, results for some of the then relatively imprecise glucose methods were falsely high in critically ill patients and for other meters consistently low. Yet, crucially, the same algorithm was used at all centers to determine the insulin infusion rates based on the glucose concentrations. The magnitude of the ensuing risk to patients with falsely high results was not apparent because no data were reported on the quality of the glucose measurements in the study.

In the aftermath of the diverging conclusions of the 2001 and 2009 studies, tight glycemic control has fallen out of

favor, but the reasons may not necessarily be justified by the research trials, Dr. Sacks says. "As soon as the NICE-SUGAR study came out, practice changed in many, many academic hospitals, and tight glycemic control was loosened substantially. The whole notion lost a lot of momentum. It's not clear whether the different centers that were involved in the multicenter study used the same meters, and it's very well known that the values from different meters vary even in the same institution." (With HbA1c, Dr. Sacks notes, because so much time and money has been spent on standardization, variations among assays are less of a problem.)

"At the time of the NICE-SUGAR study, the glucose meter accuracy was inferior to that of blood gas devices and no analytical information was provided. So we can speculate that the analytical factors might have influenced the result, but there's no way to know. I doubt we will ever know. And NICE-SUGAR is a multimillion dollar study that is, I think, undermined by the lack of information on how glucose was measured."

While the clinical laboratory community is aware of these issues, Dr. Sacks suspects most clinicians are not. "They look at a troponin result in their institution and read the literature with a completely different assay and many of them are not aware that the values are substantially different among the different methods of analysis. It's impossible to interpret a study if you don't list the manufacturer of the troponin assay." Clinicians often go to continuing education programs, he adds, but the programs deal with clinical topics. "Rarely do you have lab people talking at clinical meetings."

For the past 14 years, the AACC has tried to address the lack of clinician awareness of the need for adequate reporting. It has done so through its clinical societies collaboration committee consisting of the American Diabetes Association, Endocrine Society, and cardiology groups. Dr. Sacks, who is chair of the committee, says it's been successful but is only reaching the tip of the clinician iceberg. "This has to get out to the entire clinical community," he says.

Dr. Sacks and his coauthors submitted their article for publication to the five journals studied but were disappointed at the response. "We selected the most influential, prestigious, and highly cited clinical medical journals, the ones that publish landmark studies from which many patient diagnostic and therapeutic decisions are derived." For laboratory people, who are more aware of issues regarding analysis, the findings of their study might not be as revelatory, he believes. "We thought it would be important to publish our study in a clinical journal to get to the correct audience, so we sent it in turn to each of the five journals in the study. However, we didn't have any luck. Only one journal sent the paper out for review, and it was not accepted."

Restricted space for publication of methodological details may be a relevant obstacle to remedying the inadequate reporting as far as printed journals are concerned, Dr. Sacks concedes. But not in the digital era. "Journals are electronic. It costs a journal very, very little to add a supplement that's available on the Internet, so there's no justification for not including this information, in my opinion."

Dr. Sacks rejects the possibility that there might be ulterior motives behind a lack of reporting that biases, say, selection of patients for a study. "I think it's lack of knowledge more than an intentional attempt to undermine the system. Clearly the onus is on the journals. If the journals in their instructions to authors insist that this information be provided, then authors will provide it. But when it comes to lab tests, the journals have tended to just ignore them."



Dr. Bossuyt: "A failure to describe study methods in sufficient detail for replication renders all attempts at reproducing the original findings hazardous," he writes in his *Clinical Chemistry* editorial.

The dwindling attention at medical schools to clinical pathology or laboratory medicine and practice may be part of the problem. At the University of Amsterdam, Dr. Bossuyt writes, "medical students learn close to nothing about analytical performance and biological variability, and those who continue to be trained in clinical research methodology do not fare better."

Dr. Sacks agrees: "People who design the agenda and curricula don't perceive this as being very important, and it's not taught. Students who end up becoming clinicians are often totally unaware. A lot of clinicians view the lab as a sort of black box: A patient sample is sent to the lab, and then a 'magic number' suddenly appears in the patient's chart. I don't blame them because they focus on other things. But some understanding of the limitations of lab tests and some knowledge is essential to being a good clinician."

The Sun, et al., study recommends fuller reporting of analytical characteristics to enable investigators and others to better evaluate study results, assess the generalizability of findings, and compare and replicate results among clinical studies. As a minimal set of characteristics to report, the authors suggest three items: citation of a publication that describes the performance of the method and, if commercially available, at least the name of the product and its manufacturer; reference interval or other decision points used; and imprecision, measured during the study, at the concentrations used as decision points in the study and at non-zero limits of the reference interval. If presence or absence of a detectable or measurable concentration of a biomarker is used for decisions, the limit of detection or limit of quantification of the procedure would also be important to include. But opportunities for improvement also include simple steps such as identifying the manufacturer when a commercially available measurement procedure is used, the authors add.

Should every clinical guideline be based only on studies that include adequate reporting of analytical characteristics? Dr. Sacks suspects that is a quixotic goal. "If you take it to that extreme, you might be unable to develop adequate guidelines. If the requirement is that every single study that's evaluated in the guideline has to have reporting of analytical characteristics, then you can't do anything. But I think the information should be supplied so that one can have more confidence in the results and conclusions." The major laboratory groups would support an initiative to promote a minimal set of characteristics, in his view.

It's not fair or even reasonable to conclude that a study is flawed and should be discarded because it doesn't provide this information, Dr. Sacks cautions. "The studies may have been conducted perfectly adequately and

many probably were. And the lab analyses were probably done very well. But not being able to obtain the information, one cannot draw these conclusions.”

Nevertheless, he says, “Laboratorians should be aware of how little information is provided in clinical studies, and this may motivate them to communicate better with clinicians and educate and work with them to enhance their understanding of the role of the clinical lab.”

Dr. Bossuyt views the results of the study by Dr. Sacks and colleagues as a call for action by the community of laboratory professionals. “The scientific enterprise rests on replication,” he writes in his editorial. If invited to act as a peer reviewer, laboratory professionals “could point to the lack of details on analytical performance in reports of clinical trials.” If asked to comment on study protocols, “they could invite the principal investigators to be more specific on the laboratory tests that will be used to include and follow study participants. When they read trial reports with limited information, they could go online and leave a comment, or write a message to the editor, asking for more details. All these actions may gradually alleviate the dismal consequences of poor reporting on analytical issues in clinical research.”

Addressing the lack of transparency of biomarkers’ analytical characteristics should not be difficult, Dr. Sacks says. “An uncle of mine used to say there’s no rewind button in life. One can’t go back and redo these studies, but I think one can make changes that wouldn’t be difficult in the future.” These changes should be initiated “starting immediately,” he says, though he doesn’t see the major clinical journals stepping up to do it. “But they publish the most cited studies. And I suspect that if the major clinical journals announced this and said this is what we’re doing moving forward, then a lot of other clinical journals would follow suit.”□

Anne Paxton is a writer and attorney in Seattle.