Medical Decisionmaking: Let’s Not Forget the Physician

David L. Schriger, MD, MPH, David H. Newman, MD

From the University of California, Los Angeles, Los Angeles, CA (Schriger); and the Mt. Sinai School of Medicine, New York, NY (Newman).

A podcast for this article is available at www.annemergmed.com.

SEE RELATED ARTICLE, P. 209.


In the past 2 decades, the literature has been flooded with studies that attempt to define historical, physical, or laboratory findings that alone or in some combination can be used to risk-stratify emergency department (ED) patients. The rationale for this work is that such risk-stratification tools will identify low-risk patients who do not require additional evaluation during their ED visit. Because the elements of the tool (eg, history, physical, laboratory testing) are typically easier to acquire than the alternative action (eg, computer imaging), the tool will ostensibly decrease use, reduce costs, and avoid adverse effects of the alternate test (eg, consequences of ionizing radiation).

In this issue, Zongo et al present their work on S100-B, a marker they posit may identify patients at such low risk for clinically important head injury that cranial computed tomography (CT) can be avoided. Their article is similar to many other risk-stratification tool efforts in that it systematically discounts the physician’s capacity to stratify patients without such heuristics.

Research that fails to test the decision rule against physician judgment is implicitly predicated on 3 assumptions: (1) in the absence of the tool, the physician will order the alternate test or therapy; (2) the prediction tool will be more accurate than physician judgment; and (3) the tool neutralizes subjectivity either by posing defined clinical criteria or by introducing biomarkers as the arbiter of further resource use.

The first of these assumptions is sometimes true. Some tests are reflexively ordered for the majority of patients with a given chief complaint. For most tests, however, practice is highly variable both among individuals and among institutions. We certainly should not assume, however, that in the absence of a decision rule the alternate test will be performed. For example, the use of head CT in minor trauma is highly variable and constantly changing according to such factors as availability of CT, societal beliefs about benefits and harms, and reimbursement—factors that often have little to do with the medical value of the test.

The second assumption, that physicians make less accurate decisions without a rule, is systematically unstudied and lacks face validity. Consider an analogy: there is often vocal disapproval in the medical community when a new drug or device is studied in a randomized trial comparing it with placebo or a suboptimal alternative. Critics appropriately point out that these “controls” do not accurately represent clinical practice in the absence of the new drug. Any new drug, it is argued, should be compared with the current best treatment. Presuming this logic to be sound, there should be equal distrust of any risk-stratification tool that has been evaluated without comparison with its natural alternative, physician judgment.

With this comparative standard, risk-stratification tools such as clinical decision aids, serum markers, and clinical guidelines have often been profoundly disappointing. This is particularly true in the case of tools such as S100-B, putatively designed to reduce the need for imaging. In studies of D-dimer for pulmonary embolism, a validated clinical decision aid for minor head injury, clinical guidelines for imaging in low back pain, and a validated decision aid to reduce ankle imaging, risk-stratification tools have led to increases in the use of imaging resources. Thus, there is little empiric reason to believe that S100-B or any other biomarker should outperform a trained physician in any metric and good reason to doubt this proposition. Apart from human chorionic gonadotropin to detect very early pregnancy in fertile, sexually active women, few biomarkers or clinical decision aids have been demonstrated to be superior to physician judgment.

Finally, the third assumption—that a decision rule objectifies judgment—is erroneous. Humans introduce subjectivity in many ways. When a patient reports that he or she has a headache, substernal chest pain, and joint pains and was in a motor vehicle crash 2 days earlier, the triage nurse must choose a single chief complaint, one that may define which decision aid gets applied; when a physician is told that a headache “came on over 5 minutes and is the worst of my life,” he or she may or may not classify this as a “thunderclap” headache; providers must decide how hard to press when eliciting “tenderness of the lateral malleolus”; clinicians must similarly decide which patients generate adequate suspicion to trigger serum D-dimer testing. As these examples demonstrate, risk-stratification tools do not eliminate subjectivity; they merely shift it from one domain (eg, does this patient need this test?) to another (eg, is this patient eligible for this decision rule?).

The most common existing research model for clinical prediction instruments (of which risk-stratification tools are one type) has traditionally included 3 phases: derivation, validation, and impact analysis. In the derivation phase, a tool is...
developed according to projected sensitivity and specificity. Validation studies then test the tool to ensure that the test characteristics witnessed in the development phase persist when the rule is applied clinically. In the final step, impact analysis studies compare the use of the tool to nonuse of the tool, most often with metrics such as use of imaging or clinical outcomes. This model is flawed. Why should comparison to physician judgment be the final phase? Should not the first step be to evaluate whether the rule or biomarker offers anything beyond what can be achieved by an unaided clinician? Future risk-stratification tool studies should turn the model on its head by testing unstructured physician judgment in the derivation and validation phases. In the case of S100-B, studies should include a physician judgment limb that could easily be conducted by asking the physician before the test results are known whether he or she believes that the patient is going to have (a) an abnormal CT result and (b) an abnormal CT result that results in treatment. This small addition would put physician judgment, our current practice standard, back in to the equation before widespread use of the tool. And when decision rule evaluations include a physician judgment limb, it is likely that we will begin to see results suggesting that, for many clinical questions, physician judgment is as good as or better than biomarkers or decision aids. Studies showing that physician judgment, although imperfect, is equivalent to or better than risk-stratification tools will empower physicians and should inform society that physicians who make such judgments with due care should be no more or less responsible for rare bad outcomes than physicians who use biomarkers or “validated” decision aids.

Institutional review boards and funding agencies should not approve studies of risk stratification that do not have a physician judgment limb because the absence of such a limb institutionalizes the false notion that biomarker testing and other pseudo-objective forms of medical practice are superior to physician evaluation alone. Instead of accepting such a notion, other pseudo-objective forms of medical practice are superior to institutionalizes the false notion that biomarker testing and judgment limb because the absence of such a limb

Dr. Hollander was the supervising editor on this article. Dr. Schriger did not participate in the editorial review or decision to publish this article.

**Publication date:** Available online September 23, 2011.

**Address for correspondence:** David L. Schriger, MD, MPH, E-mail schriger@ucla.edu.

**REFERENCES**


**Supervising editor:** Judd E. Hollander, MD

**Funding and support:** By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). Dr. Schriger is supported in part by an unrestricted grant from the Korein Foundation.