REFERENCES


13 Dantzer DR. The gastrointestinal tract: the canary of the body? J Amer Med Assoc 1993; 270:1247-1248


22 Russell JA. Gastric tonometry: does it work? Intensive Care Med 1997; 23:3-6


24 Schlichtig R, Mehta N, Gayowski TJP. Tissue-arterial Pco2 difference is a better marker of ischemia than intramural pH (pHr). J Crit Care 1996; 11:51-56


Methodologic Standards for the Need to Evaluate the Standards

Contemporary physicians are extremely dependent upon laboratory testing to help guide clinical diagnosis and decision-making. By extension, we are equally dependent upon published evaluations of diagnostic tests to enable us to draw appropriate conclusions concerning the results of these tests. If,

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21828/ on 06/27/2017
as occasionally occurs, we are misled by initial incorrect reports of the discriminative value of a diagnostic test, considerable harm and expense may be inflicted upon our patients.

There are many reasons why a test might not perform as well “in the real world” as it does when initially evaluated by researchers. The most obvious would be that the test is technically difficult to perform, and can therefore only be done by extremely skilled technicians or interpreted appropriately by clinicians with special expertise. A less obvious but often more important reason is that the studied population is different from the subject population to which the test is eventually applied. In epidemiologic jargon, selection bias may weaken the external validity of a study that attempts to evaluate the usefulness of the diagnostic test.

The utility of a test depends upon characteristics of the population in which it is used, including prevalence and relative severity of disease. The mathematical formula that defines the positive or negative predictive value of a test as the probability that a patient does or does not have a disease depending upon the results of that test is influenced by the underlying disease prevalence in the population. In other words, the positive predictive value of a 10 mm purified protein derivative (PPD) reaction will be higher among contacts of a tuberculosis patient than among middle-class school children. Sensitivity (the probability a test will be positive in the diseased) and specificity (the probability that the test will be negative in the undiseased) are arithmetically independent of disease prevalence: these test characteristics will be the same for the PPD in high and low prevalence populations. On the other hand, if characteristics that influence the ability of a test to discriminate disease from nondisease are different in two populations, this will affect sensitivity and specificity. A 10 mm PPD is less sensitive for tuberculosis testing in a population of patients with AIDS, compared to an HIV-negative group with the same prevalence of disease, because immunodeficiency makes a 10-mm diameter of induration less likely in a patient with tuberculosis. More importantly for many new diagnostic tests, a group of patients participating in initial testing might have clearly defined and established disease, easily picked up by the new diagnostic test in question. When used in patients with early or mild clinical disease in the general population, the sensitivity might be far less.

Another well-defined source of bias has been termed verification bias, which occurs when not all patients receive the “gold standard” evaluation to rule out or rule in disease. Suppose a new test for pulmonary embolus (PE) is being evaluated. Patients with a positive test will have that test confirmed with a pulmonary angiogram. Those with a negative test get a lung scan, and then an angiogram if the lung scan is “high probability.” In other words, those with a positive test are more likely to receive the definitive evaluation than those with a negative test; as a result, some patients with a negative test will have an undiagnosed PE, but no patients with a positive test will have an undiagnosed PE. This would result in an underestimate of sensitivity and an overestimate of specificity. If only those patients who had angiography are included in the analysis, the estimate of specificity will be spuriously decreased. Similarly, if a study of a new test is done by evaluating it in patients who have had their PE-confirmed by angiography, but not all patients have their PE confirmed by angiography, then the discriminative features of the new test will be incorrectly estimated.

In this issue of CHEST (see page 877) Heffner and colleagues turn a critical eye on publications that have evaluated various diagnostic tests in pulmonary medicine. In doing so, they follow in the footsteps of several other authors who have applied a list of standard criteria to publications examining diagnostic tests in general medicine and pediatrics and pointed out the deficiencies therein. The findings in this study are similar to what has been previously found; 39 of the 41 evaluated publications were methodologically flawed, to a greater or lesser degree.

This is an important finding, and needs to be taken seriously because of its implications for the validity of our techniques for evaluating new diagnostic technologies. As discussed above, methodologic problems can invalidate estimates of test precision, leading to results that may mislead clinicians who attempt to use them. On the other hand, despite their deficiencies, many (if not most) of the evaluated reports provided accurate results concerning the diagnostic tests they reported on. If adherence to “methodological correctness” were evaluated as a predictor of validity of a publication, its specificity would be quite poor.

Heffner and colleagues list of methodologic standards is based upon ideal epidemiologic pedagogy, but its ultimate value needs to be demonstrated. Granted that many, if not most studies, are flawed in some way, the important question is how to weed out those with fatal flaws. Certain methodologic errors can lead to completely misleading results, and others are undesirable but not ruinous, so it is not a simple matter of counting criteria and assuming that validity is proportional to the number fulfilled. I would suggest that the criteria outlined by workers such as Heffner and colleagues need to be evaluated themselves. It would be useful to systematically compare
reports of diagnostic tests that were subsequently shown to be incorrect to reports whose validity stood the test of time, and determine the methodologic problems most likely to cause a misleading conclusion. This has been done previously by several authors on an individual case basis.\textsuperscript{3,5} What I am proposing is analogous to outcomes research as applied to clinical problems: theoretical analysis is necessary but insufficient, we need to establish what actually works.\textsuperscript{8} Proponents of evidence-based medicine assert that expert opinion needs to be supported by concrete proof: this paradigm needs to be applied to clinical epidemiology as well.

Meanwhile, attention certainly needs to be paid to the caveats raised by studies such as that of Heffner and colleagues, especially by those responsible for material published in the peer-reviewed medical literature. While it is ultimately up to clinicians to know how to apply methodologic standards to the articles they read, the editors and reviewers of medical journals can play a significant role in improving the quality of manuscripts accepted for publication. In some cases, methodologic flaws are insurmountable, the study results untrustworthy, and the manuscripts should simply not be published. In other cases, a study may be methodologically sound, but the authors need to reanalyze or present their data in a more useful and reproducible way. Finally, in some cases a study may be worth reporting even though there are problems intrinsic to its methodology; in that case, an honest discussion of the research's weaknesses and a warning that the data needs careful interpretation should feature prominently in the presentation. That warning should be included in the abstract as well, so that even the most casual readers are made aware of the study's limitations.

Michael S. Schechter, MD, FCCP
Winston-Salem, NC

Assistant Professor, Department of Pediatrics, and Director, Cystic Fibrosis Clinic, Wake Forest University School of Medicine and Brenner Children's Hospital.

REFERENCES

4 Greene RA, Begg CB. Methodology for unbiased estimation from samples of selectively verified patients. Invest Radiol 1985; 20:751-756
7 Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. JAMA 1995; 274:645-651
8 Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268:2420-2425

Weaning From Mechanical Ventilation

What Have We Learned and What Do We Still Need to Know?

Weaning from mechanical ventilation (MV) can be defined as the process of abruptly or gradually withdrawing ventilatory support from patients whose underlying cause of respiratory failure has either improved or been resolved. While MV is lifesaving, it is associated with a number of potentially serious complications.\textsuperscript{1,2} Furthermore, MV is expensive and uncomfortable for the patient. Ideally, MV should be discontinued at the earliest possible opportunity. The problem the clinician faces is how to determine when a patient is ready to resume ventilation on his or her own. Even if the patient can maintain ventilation on his or her own, the clinician must also consider whether the patient will be able to tolerate extubation. Despite a considerable amount of work devoted to this area, the decision-making process remains imperfect. Reintubation rates after discontinuation of MV range from 3 to 19\%, i.e., false-positives are relatively common.\textsuperscript{3} Assessing the incidence of false-negatives (patients continuing to receive MV who no longer require it) is more difficult. However, the self-extubation literature provides a clue. Approximately 50\% of patients who extubate themselves do not require reintubation,\textsuperscript{4,5} suggesting that the number of patients being mechanically ventilated who no longer require it is considerable.

In this issue of CHEST, (see page 886) Manthous and colleagues comprehensively review the process of liberating patients from MV, focusing on the advances made since their first such review in 1987. What important lessons have we learned since then, and what do we still need to know? Two large multicenter studies have recently examined "weaning" from MV.\textsuperscript{6,7} In these two studies, all patients