

PEDAGOGICAL TOOLS AND METHODS

# Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias

Andrew Worster, MD, MSc;\* Christopher Carpenter, MD, MSc†

SEE RELATED ARTICLE ON PAGE 131

A diagnostic test's ability to discriminate between those with and without a condition of interest (e.g., a specific disease) is best determined by comparing the test result to a reference standard that is accepted as the "truth." The evaluation of a diagnostic test's performance involves comparing its results to this reference standard,<sup>1,2</sup> and commonly reported test characteristics such as sensitivity and specificity can be determined during this kind of assessment (Fig. 1). The reference standard can take several forms. These include other diagnostic tests such as magnetic resonance imaging, a histopathology report or findings on autopsy. A reference standard may be a clinical outcome such as death or the resolution of symptoms. The reference standard may also be the results of a clinical evaluation or expert opinion. At times, the most challenging aspect of assessing a diagnostic test is identifying a reference standard that will be widely accepted.<sup>3</sup> Sometimes no acceptable reference standard exists. There is some recent work into dealing with this dilemma.<sup>4</sup>

Ideally, the reference standard and the diagnostic test under consideration are entirely independent of one another. When both the diagnostic test and the reference standard are laboratory tests, this is rarely a problem. Similarly, when the reference standard includes death, the independence of the diagnostic test and reference standard is seldom in question. However, when some clinical judgment or interpretation is required and the physician uses the diagnostic test in his or her decision-making, independence is lost.

In this issue of *CJEM*, Mater and colleagues evaluated the diagnostic test characteristics of plain radiographs in the form of a shunt series and CT scanning of the head to determine whether children have cerebrospinal shunt malfunction.<sup>5</sup> These authors chose a clinical assessment as the reference standard, and specifically defined shunt

Reference standard		
	Condition present	Condition absent
Diagnostic test positive	True positive	False positive
Diagnostic test negative	False negative	True negative

  

$$\text{Sensitivity} = \frac{\text{No. of true positives}}{\text{No. with condition present}}$$
  

$$\text{Specificity} = \frac{\text{No. of true negatives}}{\text{No. with condition absent}}$$
  

$$\text{Diagnostic accuracy} = \frac{\text{No. of true positives} + \text{no. of true negatives}}{\text{Total no. of subjects}}$$

Fig. 1. Selected diagnostic test characteristics based on a reference standard.

From the \*Division of Emergency Medicine, McMaster University, Hamilton, Ont., and the †Department of Emergency Medicine, Washington University, St. Louis, Miss.

Received: Jan. 9, 2008; accepted: Jan. 14, 2008.

*This article has not been peer reviewed.*

*CJEM* 2008;10(2):174-5

malfunction as the decision of the neurosurgeon to perform a shunt revision. Given the absence of any other generally accepted reference standard, this definition was probably reasonable and is similar to that used in other recent studies.<sup>6,7</sup> However, since the neurosurgeons undoubtedly used the results of the shunt series and CT scan of the head in their decision-making, this study is at risk for incorporation bias, one form of diagnostic test bias.<sup>8</sup> This can occur when the diagnostic test under consideration is used to determine the reference standard, or the reference standard is used to determine the results of the diagnostic test. Incorporation bias is conceptually simple. For example, we're much more likely to find signs of ischemia on an electrocardiogram after we discover an elevated troponin level. Similarly, radiologists are more likely to identify pulmonary infiltrates when the clinical information provided strongly suggests pneumonia.<sup>9</sup>

The main problem with incorporation bias is the overestimation of diagnostic accuracy.<sup>10,11</sup> In other words, the proportion of correct results is expected to be higher than one would expect if incorporation bias was not present. This, of course, would lead to overall better test characteristics, such as sensitivity. Mater and colleagues have likely introduced incorporation bias into their study. Thus, as the authors rightly point out, they have probably overestimated the sensitivity of the shunt series and head CT scans in identifying children with shunt malfunction. Notably, the authors have concluded that the test characteristics of these radiographic studies are relatively poor. For their study, incorporation bias has little impact because it is likely that the test characteristics are even worse than the authors found in their study. Since the authors conclude that the tests are relatively "bad," so to speak, it matters little that in reality the tests are even worse than they demonstrated. Unlike the study by Mater and colleagues, the situation in which incorporation bias really matters is when excellent test characteristics are found.

It is difficult to perform any research that is completely free of bias. As informed readers, it is important to understand various forms of bias that researchers may have introduced into their studies. It is equally important to understand how a particular form of bias may impact a study's results. Mater and colleagues provide an excellent example of how bias can be inherent in a study design and yet have little impact on the results of the study. Just because a study is at risk for bias being present, doesn't make the

paper "bad" or "useless." By understanding how a particular type of bias may impact a study's results, readers can avoid being biased about bias.

**Acknowledgements:** The authors would like to thank Dr. Lance Brown, Loma Linda University, Loma Linda, California, for his expert contribution to this article.

**Competing interests:** None declared.

**Keywords:** diagnostic test, reference standard, incorporation bias

## References

1. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003;138:W1-12.
2. Statement STARD. Standards for the reporting of diagnostic accuracy studies. Available: [www.stard-statement.org](http://www.stard-statement.org) (accessed 2008 Jan 11).
3. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 2003;56:1118-28.
4. Rutjes AW, Reitsma JB, Coomarasamy A, et al. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11:1-72.
5. Mater A, Shroff M, Al-Farsi S, et al. Diagnostic accuracy of shunt series and CT in the initial evaluation of cerebrospinal fluid shunt malfunction in children presenting to the emergency department. *CJEM* 2008;10:131-5.
6. Kim TY, Brown L, Stewart GM. Test characteristics of parent's visual analog scale score in predicting ventriculoperitoneal shunt malfunction in the pediatric emergency department. *Pediatr Emerg Care* 2007;23:549-52.
7. Kim TY, Stewart G, Voth M, et al. Signs and symptoms of cerebrospinal fluid shunt malfunction in the pediatric emergency department. *Pediatr Emerg Care* 2006;22:28-34.
8. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
9. Kramer MS, Roberts-Bräuer R, Williams RL. Bias and 'overcall' in interpreting chest radiographs in young febrile children. *Pediatrics* 1992;90:11-3.
10. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-76.
11. Mower WR. Evaluating bias and variability in diagnostic test reports. *Ann Emerg Med* 1999;33:85-91.

**Correspondence to:** Dr. Andrew Worster, Research Director, Department of Emergency Medicine, Hamilton Health Sciences and McMaster University, 237 Barton St. E., Hamilton ON L8N 3Z5; [worster@hhsc.ca](mailto:worster@hhsc.ca)