JOURNAL OF CLINICAL RESEARCH BEST PRACTICES

Vol. 10, No. 2, February 2014 "Happy Trials to You"

# Making Sense of Biostatistics: Using *SpPin* and *SnNout* to Evaluate Diagnostic Tests

**By Kathleen Mathieson**

Clinical trials that evaluate the effectiveness of diagnostic procedures hope to establish high values for both sensitivity and specificity. Sensitivity is the probability of a positive test, given the presence of the disease, while specificity is the probability of a negative test, given the absence of the disease.[1,2] A diagnostic test that is both sensitive and specific is thus likely to give the correct result in either the presence or the absence of the disease. In other words, it will minimize the chance of both a false negative (when a disease is present) and a false positive (when the disease is not present).

The terms "sensitivity" and "specificity" are misleading and often misinterpreted. The natural inclination is to conclude that a highly *sensitive* test is effective at identifying those *with* the disease, and a highly *specific* test is effective at identifying those *without* the disease. Counterintuitively, the reverse is true. A highly sensitive test is effective at *ruling out* the disease, while a highly specific test is effective at *ruling in* the disease. To help remember this, Sackett and colleagues[3] developed the following acronyms:

*SnNout* ("snout") is the mnemonic for: "If a test has high Sensitivity, a Negative result helps rule out the disease." In other words, if a patient actually *does have* the disease, we would expect a highly sensitive test to be positive. Therefore, when a highly sensitive test is *negative*, we can confidently *rule out* the disease.

*SpPin* ("spin") is the mnemonic for: "If a test has a high Specificity, a Positive result helps rule in the disease." In other words, if a patient actually *does not have* the disease, we would expect a highly specific test to be negative. Therefore, when a highly specific test is *positive*, we can confidently *rule in* the disease.

**Table 1. Sensitivity and Specificity**

| | Actual Disease Status | | |
|---|---|---|---|
| Test Result | (+) | (-) | Total |
| (+) | 119 | 97 | 206 |
| (-) | 1 | 118 | 119 |
| Total | 120 | 215 | 325 |
| Sensitivity=119/120=.99 Specificity=118/215=.55 | | | |

In Table 1, among the 120 patients who are known to have the disease, 119 (99%) had a positive test result (sensitivity=.99). Given the high sensitivity of this test in this sample, we would apply *SnNout* and conclude that a *negative* test result is effective for *ruling out* the disease. The highly sensitive test in Table 1 would therefore result in few false negatives.

Conversely, the test in Table 1 has poor specificity, since among the 215 patients who are known to be disease-free, only 118 (55%) had a negative test result (specificity=.55). Given the low specificity of this test in this sample, we would apply *SpPin* and conclude that a *positive* test result is *not* effective for *ruling in* the disease. The low-specificity test in Table 1 would therefore result in many false positives.

A diagnostic test ideally would demonstrate both high sensitivity and high specificity. However, there is often a trade-off — one test has high sensitivity and low specificity, while another has low sensitivity and high specificity. In many clinical contexts, one may be emphasized over the other. For example, Gross and Niedens[4] sought to validate a decision instrument to identify pelvic fractures in blunt trauma patients prior to pelvic radiography. The goal was to reduce the number of routine pelvic radiographs by predicting which patients were likely to have a pelvic fracture. In this case, because missing a pelvic fracture has serious clinical implications, the authors aimed to reduce false negatives via a highly sensitive decision instrument. In cases where false negatives have less serious implications, lower sensitivity may be acceptable, and in cases where a low rate of false positives is important, specificity may be prioritized over sensitivity. In addition, combining two or more diagnostic tests with different levels of sensitivity and/or specificity may result in better overall accuracy.[5]

**References**


1. Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM (Third Edition). New York, NY: Elsevier; 2005.
2. Goldfarb NM. Making sense of biostatistics: Sensitivity vs. specificity. Journal of Clinical Research Best Practices, 2011; 7(4). Retrieved from https://firstclinical.com/journal/contents.html?when=April+2011&layout=
3. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine (Second Edition). Boston: Little, Brown, & Co; 1991.
4. Gross, EA and Niedens BA. Validation of a decision instrument to limit pelvic radiography in blunt trauma. The Journal of Emergency Medicine, 2005; 28(3); 263-266.
5. Pepe MS & Thompson ML. Combining diagnostic tests results to increase accuracy. Biostatistics, 2000; 1(2); 123-140.


**Author**


Kathleen Mathieson, PhD, CIP, is an Associate Professor in the Doctor of Health Sciences Program at A.T. Still University (ATSU) in Mesa, Arizona. She is also a Vice-Chair of the ATSU-Arizona Institutional Review Board and a Still Research Institute Scientist. Contact her at 1.602.573.6547 or kmathieson@atsu.edu.