Interpretation of dichotomous outcomes

sensitivity, specificity, likelihood ratios, and pre-test and post-test probability

Kent, Peter; Hancock, Mark J

Published in:
Journal of Physiotherapy

DOI:
10.1016/j.jphys.2016.08.008

Publication date:
2016

Document version:
Final published version

Journal of Physiotherapy 62 (2016) 231–233

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Appraisal Research Note

Interpretation of dichotomous outcomes: sensitivity, specificity, likelihood ratios, and pre-test and post-test probability

Introduction

This is the second Research Note in a two-part series on the interpretation of statistical methods used to analyse dichotomous outcomes. The first paper covered risk, odds, risk ratios, odds ratios, and number needed to treat. This paper focuses on sensitivity, specificity, likelihood ratios, and pre-test and post-test probability. These additional measures are calculated from 2 x 2 contingency tables, as were the risk ratios and odds ratios previously covered. An example of a contingency table is shown in Box 1.

Sensitivity and specificity

Sensitivity and specificity are measures that describe how well a clinical test performs. They describe the diagnostic performance of the test in a group of patients by comparing for each patient their result on the test with whether they actually have the condition of interest (diagnosis or outcome), as indicated by a dichotomous reference standard (such as a validated questionnaire, laboratory test, imaging result or clinical outcome such as death).

Sensitivity is the proportion of people who actually have the condition of interest who are correctly identified by the test with a positive result. For example, imagine a hypothetical study (shown in Box 1) where 500 patients with low back pain and leg pain were classified as having radiculopathy if they were positive on a reference standard of concordant myotomal weakness, dermatomal sensory deficits, diminished reflexes and MRI findings of nerve root compromise. These 500 patients were also tested with the Slump Test. Such a study can assess how well the Slump Test identifies those who have radiculopathy according to the reference test. Here, the Slump Test is called the 'index test' (ie, the test that is having its diagnostic performance assessed). In this hypothetical study, 200 patients actually had radiculopathy, according to the reference test, and 150 of them were also positive on the Slump Test (true positives). The sensitivity of the Slump Test is calculated as the proportion of patients who actually had radiculopathy on the reference test and who were also correctly identified as Slump Test positive: 150/200 = 75%. This means that three out of four patients who had radiculopathy were positive on the Slump Test.

Specificity is the proportion of people who do not have the condition of interest and who are correctly identified by the test. In this hypothetical study, 300 patients did not actually have radiculopathy, according to the reference test, and 250 of them were also negative on the Slump Test (true negatives). The specificity of the Slump Test is calculated as the proportion of patients who did not have radiculopathy according to the reference test and who were correctly identified as Slump Test negative: 250/300 = 83%. This means that approximately four out of five patients who did not have radiculopathy were negative on the Slump Test.

Likelihood ratios

Likelihood ratios are another way of calculating the accuracy of the index test. Likelihood ratios determine how much more likely a particular test result is among people who have the condition of interest than it is among people who don't have the condition.[1] There is a likelihood ratio for people who are positive on a test (a positive likelihood ratio) and a different likelihood ratio for people who are negative on a test (a negative likelihood ratio). Likelihood ratios summarise the information contained in both sensitivity and specificity.[2]

A positive likelihood ratio is a measure of how much more likely a positive test result is among people who have the condition of interest than it is among people who do not have the condition of interest.[1] It is calculated as:

$$sensitivity/(1-specificity)$$

which is equal to:

$$\frac{true\ positives/(true\ positives+false\ negatives)}{1-(true\ negatives/(true\ negatives+false\ positives))}$$

In the hypothetical example above, the positive likelihood ratio would be:

$$(150/200)/(1-(250/300))=4.5$$

which indicates that a positive Slump Test is 4.5 times more likely among people who actually have radiculopathy than among those who do not have radiculopathy.

A negative likelihood ratio is a measure of how much more likely a negative test result is among people who have the condition of interest than it is among people who do not have the condition of interest.[1] It is calculated as:

$$(1-sensitivity)/specificity$$

which is equal to:

$$\frac{1-(true\ positives/(true\ positives+false\ negatives))}{true\ negatives/(true\ negatives+false\ positives)}$$

In the hypothetical example above, the negative likelihood ratio would be:

$$(1-(150/200))/(250/300)=0.3$$

which indicates that a negative Slump Test is about one-third as likely among people who actually have radiculopathy than among those who do not have radiculopathy.

Likelihood ratios of 1 indicate that the test is uninformative, whereas likelihood ratios of much more than 1 or closer to 0 indicate that the test is informative.

Likelihood ratios have a number of strengths. One strength is that they can be combined with the pre-test probability to calculate the post-test probability of the outcome, as discussed in the next section. Another strength of likelihood ratios is that they

http://dx.doi.org/10.1016/j.jphys.2016.08.008

Box 1. Calculation of sensitivity, specificity, predictive values, likelihood ratios, pre-test and post-test probability using hypothetical data.

In a cross-sectional study involving 500 patients, the diagnostic utility of the 'Slump Test' was assessed by comparing it with a reference test for radiculopathy.

	Had radiculopathy	Had no radiculopathy	Row total
Positive Slump Test	150 (true positives)	50 (false positives)	200
Negative Slump Test	50 (false negatives)	250 (true negatives)	300
Column total	200	300	500

Sensitivity

= the probability that the people who have the condition[a] will test positive
= true positives / (true positives + false negatives)
= 150 / (150 + 50) = 75% or 0.75

Three out of four patients who had radiculopathy were positive on the Slump Test.

Specificity

= the probability that the people who do not have the condition[a] will test negative
= true negatives / (true negatives + false positives)
= 250 / (250 + 50) = 83% or 0.83

Approximately four out of five patients who did not have radiculopathy were negative on the Slump Test.

Pre-test probability

= the risk or population prevalence
= 200 / 500 = 40% or 0.40

Two out of five patients in the study had radiculopathy.

Pre-test odds

= pre-test probability / (1 – pre-test probability)
= 0.40 / (1 – 0.4) = 67% or 0.67

Positive likelihood ratio

= how much more likely a positive test finding is in people who have the condition[a] than it is in people who don't have the condition[a]
= sensitivity / (1 – specificity)
= 0.75 / (1 – 0.83) = 4.5

A positive Slump Test is four and a half times more likely among patients with radiculopathy than among patients without radiculopathy.

Post-test probability (positive test result)

= (pre-test odds × positive likelihood ratio) / (1 + (pre-test odds × positive likelihood ratio))
= (0.67 × 4.5) / (1 + (0.67 × 4.5))
= 0.75

If a patient in this population has a positive Slump Test, the pre-test estimate of the probability of radiculopathy (that is, 0.40) can be revised up to 0.75.

Negative likelihood ratio

= how much more likely a negative test finding is in people who have the condition[a] than it is in people who don't have the condition[a]
= (1 – sensitivity) / specificity
= (1 – 0.75) / 0.83 = 0.3

A negative Slump Test is about a third as likely among patients with radiculopathy than among patients without radiculopathy.

Post-test probability (negative test result)

= (pre-test odds × negative likelihood ratio) / (1 + (pre-test odds × negative likelihood ratio))
= (0.67 × 0.3) / (1 + (0.67 × 0.3))
= 0.17

If a patient has a negative Slump Test, the pre-test estimate of the probability of radiculopathy (that is, 0.40) can be revised down to 0.17.

Italic text presents the interpretation of the statistic from the hypothetical study in sentence format.
[a] as determined by the reference standard test

are applicable in populations in which the condition of interest may have a different prevalence to the population in which the likelihood ratio was calculated.[1] A further strength is that they can also be used with outcomes that are not dichotomous,[1] although such use is beyond the scope of this paper.

Pre-test and post-test probability

It is useful to know the difference between the pre-test and post-test probabilities of a certain outcome because this provides a measure of the value of a clinical test. A large difference between the post-test probability of the outcome for people with a positive test result and the post-test probability for people with a negative test result is one measure of a clinically useful test. In the context of a diagnostic study, the pre-test probability is simply the probability of the diagnosis before the test is performed (continuing the hypothetical example, the probability of radiculopathy in a particular population before the Slump Test is performed).

While pre-test and post-test probabilities are easy to interpret clinically, the journey between them is somewhat convoluted, as the pre-test probability first needs to be converted into a pre-test odds, which is then multiplied by the likelihood ratio to calculate the post-test odds, which is then converted into the post-test probability. Thankfully, simple tools are available so that clinicians do not need to do the math. Tools such as nomograms and apps are available that require only the pre-test probability and the likelihood ratio to be input and they will provide the post-test probability (eg, http://www.sample-size.net/post-probability-calculator-test-new/).

Box 1 presents a worked example of the calculation of a likelihood ratio and how this is used to determine the post-test probability from the pre-test probability. In that hypothetical example, the pre-test probability is 40%. This indicates that in the absence of Slump Test results or any other diagnostic information, the best estimate is that two out of five patients in the sample have radiculopathy. The post-test probability for those with a positive Slump Test is 75% (three out of four patients with a positive Slump Test have radiculopathy) and for those with a negative Slump Test is 17% (approximately one out of five patients with negative Slump Test has radiculopathy).

The study by Downie et al[3] is a real example of using likelihood ratios to move between the pre-test probability and post-test probability. This study investigated the change in probability of having a vertebral fracture if a person presenting with low back pain had a history of prolonged corticosteroid use. They reported a likelihood ratio of 48.5. Given a pre-test probability of a person presenting for care with back pain having a fracture being 1% (based on previous research), the post-test probability of fracture in someone who has been exposed to prolonged corticosteroid use would be calculated to be 33%.

When is a test clinically useful?

The clinical usefulness of a test cannot be determined by either the sensitivity or specificity of the test alone; that requires consideration of both these performance measures. It is commonly believed that when using a test with a high specificity, a positive test result is effective at ruling in a condition (SpPIN), and that when using a test with high sensitivity, a negative test result is effective at ruling out a condition (SnNOUT).[4] While this seems intuitive, unfortunately it is not that simple[5–8] because this does not work in all scenarios. For example, if a hypothetical test had a specificity of 95% (true negatives = 95, false positives = 5) and a sensitivity of 5% (true positives = 5, false negatives = 95), then the post-test probability would be the same as the pre-test probability, so despite the test having high specificity, it added nothing to ruling the condition in. A similar scenario would apply to a test with a very high sensitivity but very low specificity. As the calculation of likelihood ratios includes both sensitivity and specificity, they are more helpful for determining when a test is clinically useful. Even though likelihood ratios might not be intuitive for clinicians, the post-test probabilities that can be calculated using likelihood ratios are easy to interpret and straightforward to communicate to patients.

As a rough guide to interpretation, positive likelihood ratios above 10 are considered to provide strong evidence to rule in a diagnosis, whereas those between 5 to 10 provide moderate evidence, and those between 2 and 5 provide weak evidence.[9] Conversely, negative likelihood ratios below 0.1 are considered to provide strong evidence to rule out a diagnosis, whereas those between 0.1 and 0.2 provide moderate evidence, and those between 0.2 and 0.5 provide weak evidence.[9] However, the impact of a likelihood ratio is very dependent on the baseline probability. For example, a likelihood ratio of 10 would result in a post-test probability of 71% if the pre-test probability were 20%, while the same likelihood ratio would result in a post-test probability of 17% if the pre-test probability were 2%. Therefore, likelihood ratios should always be interpreted in the context of the pre-test probability of the outcome. A range of factors will influence how high the post-test probability for a positive test needs to be before the test is clinically useful (or how low for a negative test). These factors include: the consequence of the decision; whether the test is to be used on its own or in combination with other tests; the ease of performing the test; and the cost and reliability of the test.

Summary

This paper has introduced and explained sensitivity, specificity, likelihood ratios, pre-test and post-test probability, in the context of a diagnostic test. A good understanding of these terms will enable readers of clinical studies to ensure they correctly interpret the clinical importance of the findings reported.

Ethics approval: Nil.

Competing interests: Nil.

Source(s) of support: Nil.

Acknowledgements: Nil.

Competing interests: Nil.

Provenance: Not invited. Peer reviewed.

Peter Kent[a,b] and Mark J Hancock[c]

[a]*School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia*

[b]*Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Denmark*

[c]*Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia*

References

1. Grimes DA, Schulz KF. *Lancet.* 2005;365(9469):1500–1505.
2. Dujardin B, et al. *Eur J Epidemiol.* 1994;10:29–36.
3. Downie A, et al. *BMJ (Clin Res Ed).* 2013;347:f7095.
4. Sackett D, et al. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Boston, MA: Little Brown; 1992.
5. Baron JA. *Med Decis Making.* 1994;14:107.
6. Boyko EJ. *Med Decis Making.* 1994;14:175–179.
7. Hegedus EJ, Stern B. *J Man Manip Ther.* 2009;17:E1–E5.
8. Pewsner D, et al. *BMJ BMJ (Clin Res Ed).* 2004;329(7459):209–213.
9. Jaeschke R, et al. *JAMA.* 1994;271:703–707.