Interpreting diagnostic tests with continuous results and no gold standard

A common scenario explained using the tuberculin skin test

Dobler, Claudia C.; Murad, M. Hassan

Published in:
BMJ Evidence-Based Medicine

DOI:
10.1136/ebmed-2017-110825

Licence:
Other

Interpreting diagnostic tests with continuous results and no gold standard: A common scenario explained using the tuberculin skin test

Claudia C. Dobler[1], M. Hassan Murad[1]

1) Evidence-based Practice Center, Mayo Clinic, Rochester MN, USA

Correspondence to

Dr. Claudia C Dobler

Evidence-based Practice Center

Mayo Clinic

Rochester MN 55905

USA

dobler.claudia@mayo.edu

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

ABSTRACT

Practitioners of evidence-based medicine commonly encounter diagnostic tests with continuous results and no gold standard. In contrast, the traditional critical appraisal teachings assume a binary test (2x2 table) with a gold standard. In this guide, we use the example of the tuberculin skin test (TST) to illustrate a simple approach facilitated by using stratum-specific likelihood ratios and odds of developing future patient-important events. This approach can aid practitioners in the interpretation and application of diagnostic tests to patient care.

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

A very common scenario that practitioners of evidence-based medicine (EBM) face is interpreting diagnostic tests with continuous results (e.g. size of a tuberculin skin test [TST] reaction, blood glucose level, prostate-specific antigen level in the blood). In fact, it is rare to have a diagnostic test with binary results (sick/healthy). It is also common to face tests without a gold standard.

The typical approach taught in EBM workshops and books is to use a cut off or threshold level, above which the presence of a certain condition is likely (or below which the presence of a condition is unlikely). Using this cut off, the results are dichotomized to generate a 2x2 table and likelihood ratios.[1] Practitioners use the likelihood ratios (from the literature) to determine the post-test probability of their patient as to whether they truly has (or does not have) a certain condition. The dilemma arises when test results are well above the cut off point; which should intuitively make a clinician more confident that a condition is truly present (compared to a test result just above the cutoff point). This is, however, not accounted for in an approach where the tests results are dichotomized.

In this guide, we address these two common issues using the example of TST results. We demonstrate the use of multiple cut offs (also referred to as multilevel likelihood ratios or stratum-specific likelihood ratios) to account for these intuitive differences in test results. We also address the frequent scenario of a diagnostic test that can be evaluated in absence of a gold standard by linking the test result to a future patient-important outcome.

Clinical scenario

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

You are a general practitioner working in an area with a large proportion of migrants from Southeast Asia. The incidence of tuberculosis (TB) in this area is 100 per 100,000 people per year. Earlier today you saw a 47-year old Vietnamese woman, who was diagnosed with rheumatoid arthritis and is now considered for tumour necrosis factor (TNF)-alpha inhibitor treatment. Because of the increased risk of developing an infection once she receives TNF-alpha inhibitor treatment, she had a TST to screen for latent (dormant) TB infection (LTBI). Before seeing the test result, you think that your patient's current risk of developing TB is likely consistent with the average risk in her community and is therefore 0.1% per year (100/100,000) or 2% for the next twenty years.

When you read the TST reaction 72 hours later it shows an induration measuring 20mm in diameter (TST size ≥10mm is considered a positive reaction in this patient). In the literature you find that the positive likelihood ratio (for the risk of developing TB) was 1.86 (95% CI 1.56–2.21) for the risk of developing TB within 4 years in contacts of patients with active TB with a positive TST (≥10mm).[2] You are aware that this is only an approximate estimate because the likelihood ratio for the risk of developing TB might be higher in recent contacts of a TB patient than in your patient, in whom time of TB infection is unknown. On the other hand, the (cumulative) risk of developing TB will increase with a longer observation period beyond 4 years. Using Fagans' nomogram for calculating post-test probabilities[3] or an online calculator, you determine that the prior probability of the patient to develop TB (in twenty years) of 2% has increased to a post-test probability of 3.8% because of the positive test. Intuitively you feel that the probability of developing TB in your patient with a TST reaction of 20mm should be higher

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

than in a person with a TST of 10mm. You wonder whether there is an approach where the likelihood ratio takes into account the "strength" of a positive test result.

Using multiple cut offs: stratum-specific likelihood ratios

While describing the performance of a diagnostic test using sensitivity and specificity is based on dichotomizing outcomes, computation of likelihood ratios does not require that test results are dichotomized.[4] Using stratum-specific likelihood ratios can give a more precise likelihood estimate (as the likelihood ratio is not diluted across several stratums) and thus a more personalized estimate for an individual patient with a specific test result.[5] A study of 1335 contacts of patients with active TB in Spain described stratified positive and likelihood ratios for different TST reactions. For a TST measuring ≥15mm the positive likelihood ratio was 3.59 (95% CI 2.59–4.98).

As likelihood ratios are independent of pre-test probability and prevalence, the likelihood ratio can be applied to the patient in our clinical scenario, although it was derived from a cohort that probably had a higher pre-test probability of being infected with TB because of contact with a patient with active TB. The post-test probability to develop TB in the next 20 years is thus 7.3% (pre-test odds x positive likelihood ratio) in our patient, and as such significantly higher than the post-test probability obtained with the likelihood ratio from the dichotomized test interpretation, confirming the clinician's intuition.

Evaluation of diagnostic tests when there is no gold standard

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

There is no gold standard to ascertain latent TB infection (LTBI) in a person, such as a biopsy or microbiology results, which, if, positive for *Mycobacterium tuberculosis,* by definition is indicative of active TB disease. All existing tests for LTBI --the TST and gamma interferon (IFN-γ) release assays (IGRAs)-- are indirect tests which show that the host has been sensitized to TB antigens.[6] Hence, different methods have been used to evaluate tests for LTBI (TST or the newer interferon-gamma-release assays) including the use of surrogate reference standards such as a history of contact with a patient with infectious TB, a history of TB disease, evidence of old/inactive TB on chest x-ray or a combination of these.[7] In absence of a gold standard for diagnosing LTBI, performance of the TST can also be evaluated by linking it to the development of active TB, a (future) patient-important event,[8] as in our clinical scenario.

To link the TST results to a future outcome, we need to estimate the risk of developing TB in this individual. Such risk depends on two factors: 1) the probability that true LTBI is present (based on the pretest probability and the positive likelihood ratio), and 2) the risk of progression from LTBI to active TB (Figure 1). Factors that have to be taken into account to determine the risk of progression from LTBI to active TB include the time when infection likely has occurred (as the risk for progression to active TB decreases with time after infection[9]) and medical conditions that increase the risk of progression to active TB. A number of conditions such as HIV infection,[10] diabetes,[11] renal failure[12] and being on TNF-alpha inhibitor treatment,[13] have been shown to be associated with an increased risk of TB.

Resolution of clinical scenario

The patient's risk of progressing from LTBI to TB is increased 1.8 to 29.3 times by taking TNF-alpha inhibitor treatment (depending on which TNF alpha inhibitor she will receive),[12] thus

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

increasing the risk of developing TB in the next 20 years to 13-100% (7.3x 1.8 and 7.3x 29.3; respectively). This high risk provides compelling evidence for the clinician to recommend treatment of LTBI to the patient.

For the particular challenge of estimating risks of developing TB based on a patients epidemiological and medical characteristics, there is also an online calculator available (http://www.tstin3d.com/en/calc.html).[14] Using this calculator, the patient in our example had an estimated annual risk of developing TB of 0.54% (11% in 20 years) and a cumulative risk of developing TB up to the age of 80 years of 18%. This risk estimate is lower than the estimate obtained with the likelihood ratio calculation, because the likelihood ratio was obtained from a population of recent TB contacts.[2] If the patient had been a recent close contact of a patient with TB, her risk of developing TB, using the online calculator, would rise to 27% over the next 2 years and 43% up to the age of 80 years. In any case, the risk of developing TB is sufficient to warrant LTBI treatment.

Conclusion

Diagnostic tests commonly have continuous results and no gold standard. Interpretation and application to patient care can be facilitated by using stratum-specific likelihood ratios and odds of developing future patient-important events.

Acknowledgments

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

CCD is supported by a fellowship of the Australian National Health and Medical Research Council (APP1123733). The fellowship sponsor had no role in manuscript design, data interpretation, or writing of the manuscript.

References

1. Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr* 2007;96(4):487-91. doi: 10.1111/j.1651-2227.2006.00179.x
2. Altet N, Dominguez J, Souza-Galvao ML, et al. Predicting the Development of Tuberculosis with the Tuberculin Skin Test and QuantiFERON Testing. *Annals of the American Thoracic Society* 2015;12(5):680-8. doi: 10.1513/AnnalsATS.201408-394OC
3. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975;293(5):257. doi: 10.1056/nejm197507312930513
4. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *Bmj* 2004;329(7458):168-9. doi: 10.1136/bmj.329.7458.168
5. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed., edited by Gordon Guyatt, MD, Drummond Rennie, MD, Maureen O. Meade, MD, and Deborah J. Cook, MD New York: McGraw-Hill Education, 2015:353-4.
6. Pai M, Denkinger CM, Kik SV, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. *Clinical microbiology reviews* 2014;27(1):3-20. doi: 10.1128/cmr.00034-13
7. Ferguson TW, Tangri N, Macdonald K, et al. The diagnostic accuracy of tests for latent tuberculosis infection in hemodialysis patients: a systematic review and meta-analysis. *Transplantation* 2015;99(5):1084-91. doi: 10.1097/tp.0000000000000451
8. 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(50):iii, ix-51.

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.

9. Ferebee SH. Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl Tuberc* 1970;26:28-106.
10. Corbett EL, Watt CJ, Walker N, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 2003;163(9):1009-21. doi: 10.1001/archinte.163.9.1009
11. Dobler CC, Flack JR, Marks GB. Risk of tuberculosis among people with diabetes mellitus: an Australian nationwide cohort study. *BMJ Open* 2012; 2: e000666.
12. Dobler CC, McDonald SP, Marks GB. Risk of tuberculosis in dialysis patients: a nationwide cohort study. *PLoS One* 2011; 6: e29563.
13. Dobler CC. Biologic agents and tuberculosis. *Microbiol Spectrum* 2016; 4: TNM17-0026-2016.
14. Menzies D, Gardiner G, Farhat M, et al. Thinking in three dimensions: a web-based algorithm to aid the interpretation of tuberculin skin test results. *Int J Tuberc Lung Dis* 2008;12(5):498-505.

This article has been accepted for publication in BMJ Evidence-Based Medicine (2017) following peer review, and the Version of Record can be accessed online at http://dx.doi.org/10.1136/ebmed-2017-110825.