

How to interpret diagnostic test results; an evidence based medicine approach

Klazien Matter-Walstra, Switzerland.

About the author:

Klazien Matter-Walstra studied human biology at the University of Groningen, Holland. Thereafter she came to Bern, Switzerland for her PhD. After a 7 year period work at the Institute of Pathology in Bern, where she introduced and managed a laboratory for immunocytochemical cancer diagnosis for exfoliative cytology. She became involved in evidence based medicine and epidemiology and changed jobs. At the moment she works for the company Mediscope where she is responsible for medical literature monitoring, teaching evidence based medicine and writing critical appraisals on medical publications. Her own homepage with more information and a full c.v. can be found at <http://home.tiscalinet.ch/kmatter>



Some thoughts about evidence based medicine

The term "evidence based medicine" (EBM) was coined at McMaster Medical School in Canada in the 1980s to label a clinical learning strategy, which people at the school had been developing for over a decade. (BMJ 1995;310:1122). EBM asks questions, finds and appraises the relevant data, and harnesses that information for everyday clinical practice. Together with clinical epidemiology, EBM provides clinicians and patients with new tools and consciousness to interpret diagnostic test and treatment results or risk and harm data. EBM is thought to be a new paradigm in medicine but it may also be seen as looking from another angle at the same problem. Like in physics, where the awareness of "probabilities" emerged and replaced the "Newtonian certainties", EBM also points at the basically probabilistic characteristics of medicine, and how to communicate them. For physicians as well as patients, probabilistic thinking is very difficult because emotionally humans can better handle certainties and find it hard to make decisions based on odds. On the other hand, most daily decisions made are based on intuitive probabilistic thinking. It is the awareness of the likelihood and the way it is communicated which introduces uncertainty. Together with some basic theoretical philosophy of science, this probabilistic thinking will be illustrated for medical diagnostic tests.

Some basics on diagnostic tests

The primary aim of a diagnostic test is to tell whether a person has (will have) a disease and what kind of disease. In principle the development of a diagnostic test starts with the observation that people with disease X manifest a feature Y which is not seen in people without the disease X. Feature Y may be found coincidentally or it may be explicitly searched for (for example bacteria, viruses, cancer cells, antibodies etc.). As soon as the feature Y is established as a characteristic of disease X, evaluation to determine the accuracy of diagnostic tests can start: people with and without the disease are tested for their expression of the feature. The following four possibilities will/can occur:

- A: people with feature Y, with disease X
- B: people with feature Y, without disease X
- C: people without feature Y, with the disease X
- D: people without feature Y, without the disease X

In EBM these four possibilities are put in a so-called "two-by-two table":

Two-by-two table

	Disease X	Healthy	Total
Test positive for Y	A	B	A+B
Test negative for Y	C	D	C+D
Total	A+C	B+D	A+B+C+D

A diagnostic test has four important characteristics:

1. The **sensitivity** of a test is defined as the proportion of truly diseased persons in the screened population who are identified as diseased by the screening test. Sensitivity is a measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test (true positive rate) :

$$A/(A+C)$$

2. The **specificity** of a test is defined as the proportion of truly non-diseased persons who are so identified by the screening test. It is a measure of the probability of correctly identifying a non-

diseased person with a screening test (true negative rate
 $D/(D+B)$)

3. The **positive predictive value (PPV)** of a test is defined as the probability that a person with a positive test result is true positive (i.e., does have the disease
 $A/(A+B)$)

4. The **negative predictive value (NPV)** of a test is defined as the probability that a person with a negative test result is true negative (i.e., does not have the disease
 $D/(C+D)$)

Gold standard, sensitivity and specificity, Likelihood-Quotient

During the development of a new diagnostic test for a new disease it may be unclear whether the feature Y is a direct marker for the disease or only an indirect marker (like an antibody directed against an still unknown contagious agent). As soon as the contagious agent (or some internal factor) is isolated and characterized the sensitivity and specificity of the diagnostic test (detecting an indirect marker) can be refined. For the example HIV, sensitivity then becomes the percentage of positive test results where HIV truly can be isolated divided by all positive test results. In this case all test results have to be checked for the presence of HIV, and HIV presence is called the gold standard. However, for many, if not most, medical diagnostic tests, no such "true" gold standard is available as the gold standard itself is a subjective observation.

For example, for developmental dysplasia of the hip (DDH) in the neonate, congenital hip dislocations in newborn children were diagnosed by manual hip manipulation looking for a "click" and "loose" hips. Later ultrasonography was introduced as the method for diagnosing DDH. The only gold standard in both tests is how many children with a positive diagnosis really developed DDH. None of the tests, however, had a true gold standard. The sensitivity of the ultrasonographic test was related to the old clinical test and thought to be much better, but the old test itself did not have a real proof. The discussion about what a gold standard really is, is a philosophic issue and may not be answered in the end. In medicine a gold standard is not a fact but a biological quality or state that comes as close as possible to the true condition or can predict the best course of a disease.

In addition, even if a gold standard is available, the biological value of such a standard can be questioned. For example, if an antibody test detects an influenza virus with a high sensitivity, this does not mean that all of the persons testing positive will also develop a flu. In relation to the gold standard "influenza virus" the test can be highly accurate; however, in predicting whether a person will get ill this may not be so. This dilemma is hardly discussed in medicine and as long as no natural history of a disease is known it can be delicate to start treating people because of a positive test result. For some cancers this phenomenon now becomes more and more discussed. One example is that of breast cancer, where the early detection of breast cancer cells with mammography results in the treatment of women who, if those breast cancer cells hadn't been diagnosed, would never have gotten the disease breast cancer in their life.

When are diagnostic tests used: positive predictive values

Sensitivity and specificity are characteristics of a test that are independent of the person who is tested. However, the probability that a test result is true is dependent on the likelihood that the disease (infection etc.) is present. This will be illustrated in the following two examples.

1. About a married couple who wants to live in Singapore

A happily married couple wants to live in Singapore for three years. Before they get their permission to stay they have to do an AIDS test. The husband tests positive with an AIDS test which has a 97% sensitivity and a 98% specificity (it will not be discussed here to what "gold standard" these two refer). How big is the chance that this man is truly positive? Should he start treatment immediately?

Before this man had the test it may be assumed (see *) that the probability that he is HIV infected is only 1 in 10.000, or 0.01% (see *) (he was married to his wife for more than 15 years and never has slept with other men or women, nor does he takes drugs).

Imagine 1.000.000 men like him. In this population 100 will be HIV positive (1 in 10.000 or 100 in 1.000.000). From the 100 HIV positive men the test will detect 97 (sensitivity). From the remaining 999.900 HIV negative men, however, the test will detect 98% as truly negative (specificity) and 2% as positive, which is 19998 persons. If all the 1.000.000 would have been tested in total there would be 20095 positive test results of which only 97 are truly positive. This means that the probability that the man in question is really positive is 0.48% (97 of 20095). In other words, the PPV is only 0.48%.

	Infected	Healthy	Total
Test positive for HIV (sensitivity 97% out of 100)	97	19998	20095
Test negative for HIV (specificity 98% out of 999900)	3	979902	979905

Total	100	999900	1.000.000
--------------	-----	--------	-----------

PPV:
 $(97/20095) * 100\% = 0.48\%$

Before this man did the test he could assume that the probability that he is positive for HIV was 0.01%, now after the test his probability is 0.48%. The question is if this does make a difference and if such a test result should be followed by certain consequences.

** Prevalence rate: The total number of all individuals who have an attribute or disease at a particular time (or during a particular period) divided by the population at risk of having the attribute or disease at this point in time. A prevalence for a single person may differ from that a certain population. In Europe it is assumed that the prevalence of HIV positive persons is about 2%, this number includes persons at high risk and those with a low risk. In a high risk population the prevalence may be 20%; in a very low risk population it may be 0%. For a single person the pre-test probability in the end is always an estimation, based on a combination of the known prevalence for the risk group a person belongs to (as far as is known at all) and the individual situation. In the above example the general prevalence of 2% must be estimated much lower according to the man's private situation and in the end always will be a guess. It has to be kept in mind, though, that prevalence for HIV infection are determined by tests that in themselves are questionable, making the guesswork even more obvious.*

2. A young heroin addicted prostitute is brought into a hospital after she collapsed on the street:

A young heroin addicted prostitute is brought into the emergency room. As a routine also an AIDS test is done. The test results are positive. Again an assumption can be made about the chance that this woman is HIV infected. Because of her way of life the chance is high and can be estimated to be 10%, or 1 out of 10. Imagine 1000 women like her, with the same test as above. The results will look like the following:

	Infected	Healthy	Total
Test positive for HIV (sensitivity 97% out of 100)	97	18	115
Test negative for HIV (specificity 98% out of 900)	3	882	885
Total	100	900	1000

PPV:
 $(97/115) * 100\% = 84.35\%$

In this case the pre-test probability is 10% and the post-test probability almost 84%. The gain of information in this case is tremendous and the test results should have consequences.

Although the man's probability of being truly HIV positive increases almost fifty times (from 0,01% to 0,48%) it is an increase from "nothing to nothing". On the other hand, in the case of the woman, the "only" eight times increase is from "something to much". It is important to state that great increases of very small probabilities end up in still small probabilities. Depending on the art of expressing numbers the value of a test may be misinterpreted. In relative numbers the test may increase the knowledge about HIV infection up to fifty times. In absolute numbers this may mean that before a test was done the chance of being positive was estimated to be 1 in 10.000; with a positive test result this chance increases to 48 in 10.000.

The man from example 1 may decide to have a second test and chooses PCR. Assume PCR has a sensitivity of 98% and a specificity of 99%. The pre-test probability now is 0.48%. With a calculation described as above the PPV ends up to be 32%. This number is still less than 50%. Before considering doing a test it has to be known if a certain PPV threshold value for incurring consequences (like starting treatment) exists. As long as this threshold value cannot be trespassed based on the pre-test probability, it may be questioned whether a test should be done at all. For HIV testing no such consent about treatment thresholds for positive tests are known.

In summary this means that a positive (or negative) test result does not at all mean that a person is truly positive (or negative). It only indicates a higher probability of being positive. How much higher this probability is depends on the pre-test probability of a person and the sensitivity and specificity of the test used. The lower the chance that somebody is HIV infected, the higher the chance that a positive test result is a false positive test result. This understanding has consequences for the settings in which a diagnostic test is used. When screening whole populations for a certain disease the pre-test probability in general may be low and such a screening will result in high numbers of false positive test results. For HIV this is the case in testing all immigrants or everybody who wants to join the army (in the USA the prevalence for HIV infection is 0.0% to 0.1% among applicants for military service) and even in testing all blood

donors.

Evidence based medicine stresses the significance of probabilities over certainties. No test result ever gives a true value but only a probability statement. Unfortunately in general a test result is looked upon as being true with all the following consequences. Even most AIDS counselors do not understand the principle of pre-test probability and PPV (**see Gigerenzer G, Hoffrage U, Ebert A. "AIDS counseling for low-risk clients." AIDS Care. 1998 Apr;10(2):197-211).**

Although probability thinking is difficult and oft confusing, EBM and this paper hopefully elevate the confusion to a higher level, and show that in medicine no black and white exist. One hopes this will make the reader more alert and critical when being tested.

References:

1. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1994;271:389-91.
2. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. BMJ 1997;315:540-3.
3. Richardson WS, Wilson MC, Williams JW, Jr, Moyer VA, Naylor CD. Users' guides to the medical literature: XXIV. How to use an article on the clinical manifestations of disease. Evidence-Based Medicine Working Group. JAMA 2000;284:869-75.
4. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2000;284:79-84.
5. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994;271:703-7.
6. Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis. Evidence-Based Medicine Working Group. JAMA 1999;281:1214-9.

The author can be contacted at:

<mailto:matter@mediscope.ch>, <http://home.tiscalinet.ch/kmatter>

top of page	contents vol.6 / no.3	main contents
-----------------------------	---------------------------------------	-------------------------------