

On the known, yet still pervasive, myth that sensitivity and specificity do not depend on patient covariates

Samuel J. Weisenthal

^aUniversity of Rochester Medical Center, 601 Elmwood Ave, Rochester, 14642, NY, USA

Abstract

There is a major probability error that is made in everyday medical diagnostic reasoning. The prevailing medical diagnosis framework lures users into making this error by defining sensitivity and specificity as independent of patient covariates. This leads to the mixing and matching of risk scores for pre-test probabilities with sensitivities and specificities from studies that do not take into account the same covariates as the risk scores. That sensitivity depends on covariates is well known, and certain authors have criticized the prevailing probabilistic medical reasoning framework for misleading its users to think otherwise. However, no text has explicitly written out the probability expressions. In this work, we write out the probability calculations, and we show the precise error in reasoning that occurs.

1. Introduction

Consider a scenario in which a patient presents with unilateral swelling of the calf. After clinical evaluation, one computes the Wells' pre-test probability (Wells et al., 1998) of deep venous thrombosis (DVT). Given the pre-test probability, one chooses whether to order an ultrasound based on the sensitivity and specificity of the ultrasound. Given the sensitivity and specificity of the ultrasound, and the pre-test probability of DVT from the Wells' score, one computes a post-test probability of DVT. Does this sound reasonable? If so, read on.

The scenario we have just described is very common. It is a standard usage of the prevailing medical diagnostic reasoning framework. The prevailing framework, however, which uses Bayes rule to obtain a post-test probability,

lures users into a trap. It leads users to assume that sensitivity and specificity do not depend on patient covariates. For example expositions of the prevailing medical diagnostic reasoning framework, which write sensitivity and specificity so that they do not depend on patient covariates, see, e.g., Ledley and Lusted (1959); Parikh et al. (2008); Trevethan (2017)). That sensitivity and specificity do depend on patient covariates has been noted by many authors Moons et al. (1997); Harrell et al. (2001); Moons and Harrell (2003); Guggenmoos-Holzmann and van Houwelingen (2000), but the prevailing framework is so pervasive that this is usually ignored in the clinical setting. The downstream consequence of the assumption that sensitivity and specificity do not depend on patient covariates is that almost all post-test probability calculations accidentally mix a pre-test probability derived from a risk score that depends on patient covariates with sensitivities and specificities from studies that do not take into account those same covariates. This mixing and matching can make the post-test probability meaningless. In this work, we write out the probability expressions behind the error, and we discuss its clinical consequences.

2. Background

2.1. Diagnostic problem

In the standard diagnostic problem, we seek to obtain a post-test probability of a disease given a test result. In the prevailing medical diagnosis framework, which depends on Bayes rule, if Dz is a binary random variable representing disease presence if it is set to 1 and disease absence if it is set to 0 and $Test$ is a binary random variable representing a positive test result if it is set to 1 and a negative test result if it is set to 0, we have that

$$\begin{aligned} p(Dz = 1|Test = 1) &= \frac{p(Test = 1, Dz = 1)}{p(Test = 1)} \\ &= \frac{p(Test = 1|Dz = 1)p(Dz = 1)}{p(Test = 1)} \\ &= \frac{(\text{sensitivity})(\text{pre-test probability})}{p(Test = 1)}, \end{aligned} \quad (1)$$

where sensitivity is $p(Test = 1|Dz = 1)$ and pre-test probability is $p(Dz = 1)$. Using the law of total probability, the denominator can be further written

in terms of sensitivity, specificity, which is formally defined as $p(Test = 0|Dz = 0)$, and pre-test probability:

$$\begin{aligned}
p(Test = 1) &= p(Test = 1, Dz = 1) + p(Test = 1, Dz = 0) \\
&= p(Test = 1|Dz = 1)p(Dz = 1) \\
&\quad + p(Test = 1|Dz = 0)p(Dz = 0) \\
&= p(Test = 1|Dz = 1)p(Dz = 1) \\
&\quad + (1 - p(Test = 0|Dz = 0))(1 - p(Dz = 1)) \\
&= (\text{sensitivity})(\text{pre-test probability}) \\
&\quad + (1 - \text{specificity})(1 - (\text{pre-test probability})).
\end{aligned}$$

In this way, one can write the standard post-test probability as a function of sensitivity, specificity, and pre-test probability.

Note that (1) only conditions on the test result, like most examples from the prevailing medical diagnosis literature (see the commonly cited overviews of post-test probability, such as Parikh et al. (2008); Trevethan (2017)). Conditioning only on the test result is not unreasonable, especially to introduce the concept of post-test probability for educational purposes, but it does not align with the true clinical objective of diagnostic reasoning. The clinician wishes to provide the most patient-specific post-test probability, and this probability is a function not only of the test result but also of the other patient covariates (i.e., other information that we have about the patient).

Unlike in the standard post-test probability examples where one conditions only on the test (shown in (1)), consider the case where one also conditions on patient covariates, X . For example, X might contain symptoms or demographic information. Conditional on $X = x$, we obtain the post-test probability

$$\begin{aligned}
p(Dz = 1|Test = 1, x) &= \frac{p(Test = 1|Dz = 1, x)p(Dz = 1|x)p(x)}{p(Test = 1|x)p(x)} \\
&= \frac{p(Test = 1|Dz = 1, x)p(Dz = 1|x)}{p(Test = 1|x)}. \tag{2}
\end{aligned}$$

Note that, as was done for the denominator of (1), the denominator of (2)

can be rewritten

$$\begin{aligned}
p(\text{Test} = 1|x) &= p(\text{Test} = 1, Dz = 1|x) + p(\text{Test} = 1, Dz = 0|x) \\
&= p(\text{Test} = 1|Dz = 1, x)p(Dz = 1|x) \\
&\quad + p(\text{Test} = 1|Dz = 0, x)p(Dz = 0|x) \\
&= p(\text{Test} = 1|Dz = 1, x)p(Dz = 1|x) \\
&\quad + (1 - p(\text{Test} = 0|Dz = 0, x))(1 - p(Dz = 1|x)) \\
&= (\text{sensitivity})(\text{pre-test probability}) \\
&\quad + (1 - \text{specificity})(1 - (\text{pre-test probability}))
\end{aligned}$$

Now, in (2), sensitivity is defined as $p(Dz = 1|\text{Test} = 1, x)$ instead of $p(Dz = 1|\text{Test} = 1)$, and likewise for specificity, $p(Dz = 0|\text{Test} = 0, x)$, and pre-test probability, $p(Dz = 1|x)$.¹ In general, in (2), the presence of the covariate, x , requires one to extend (1) to the more complex expression in (2), which has a more complex representation of sensitivity. This requirement to extend the expression, if not done properly, can lead to a particularly nefarious error in reasoning, which we will now describe.

Consider a scenario in which a patient presents with swelling in the calf. One starts with Wells' (Wells et al., 1998) for a deep venous thrombosis, which gives $p(DVT + |cancer, swelling, etc \dots)$. Then, given an US result, one consults the literature to find the sensitivity, $p(US + |DVT+)$, and specificity, $p(US - |DVT-)$, of the US (e.g., one consults (Goodacre et al., 2005)). Note that Goodacre et al. (2005) gives $p(US + |DVT+)$ instead of $p(US + |DVT+, cancer, swelling, etc \dots)$; i.e., the sensitivity and specificity values are not conditional on patient covariates. Using these sensitivity and specificity values in conjunction with Wells implicitly assumes (this is not often written out, it is just hidden within the diagnostic reasoning) that

$$\begin{aligned}
&p(DVT + |US+, cancer, swelling, etc...) \\
&= \frac{p(\mathbf{US} + |\mathbf{DVT+})p(DVT + |cancer, swelling, etc...)}{p(US+, cancer, swelling, etc...)}.
\end{aligned}$$

¹One could argue that this is overloading "sensitivity" to represent both $p(\text{Test} = 1|Dz = 1)$ and $p(\text{Test} = 1|Dz = 1, x)$, and we will discuss this below.

However, it should be

$$\begin{aligned} & p(DVT + |US+, cancer, swelling, etc...) \\ &= \frac{\mathbf{p}(US + |DVT+, \mathbf{cancer}, \mathbf{swelling}, \mathbf{etc...})p(DVT + |cancer, swelling, etc...)}{p(US+, cancer, swelling, etc...)}. \end{aligned}$$

This error is rampant in the clinical world, since most post-test probability calculations are done informally (the expressions are not written out).

In general, the error occurs when one defines a post-test probability that implicitly assumes

$$p(Dz = 1|Test = 1, x) = \frac{\mathbf{p}(\mathbf{Test} = \mathbf{1}|\mathbf{Dz} = \mathbf{1})p(Dz = 1|x)}{p(Test = 1, x)},$$

where the denominator is assumed to be written as

$$\begin{aligned} p(Test = 1, x) &= p(Test = 1|Dz = 1)p(Dz = 1|x) \\ &\quad + p(Test = 1|Dz = 0)p(Dz = 0|x). \end{aligned}$$

The last two displays only hold if $Test \perp X$, in which case $p(Test = 1|Dz = 1)$ truly equals $p(Test = 1|Dz = 1, x)$. Assuming $Test \perp X$, however is an error. It is unlikely that US is independent of the severity of the DVT, which is taken into account in the Wells' score based on the tenderness of the leg and extent of the swelling. That this independence between test and covariates does not hold for other medical conditions has been shown in many studies; see, for example, Moons et al. (1997); Hlatky et al. (1984); Prince-Guerra et al. (2021) or the other examples listed in (Moons and Harrell, 2003; Harrell et al., 2001; Guggenmoos-Holzmann and van Houwelingen, 2000)).

The prevailing diagnostic framework, however, does not appear to claim that the test does not depend on the patient covariates. Instead, the prevailing diagnostic framework lures users into making this assumption by naming the special case, $p(Test = 1|Dz = 1)$, as sensitivity.² Note, for example, that in virtually all expositions of the prevailing medical diagnosis framework see e.g. (Parikh et al., 2008; Trevethan, 2017) the authors define

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}},$$

²More generally, the prevailing framework causes errors by renaming probabilities, which are better left unnamed.

which is $p(\text{Test} = 1|Dz = 1)$.

But $p(\text{Test} = 1|Dz = 1)$ is a special case, because sensitivity can be more broadly defined as $p(\text{Test} = 1|Dz = 1, \text{covariates})$. If sensitivity truly depends on patient covariates, only the latter is useful to obtain the true clinical desirable, $p(Dz = 1|\text{Test} = 1, \text{covariates})$. In a sense, it is as if the post-test probability is a vehicle and sensitivity is a wheel. The prevailing framework defines a wheel to be a stone cylinder, and this has generally led the majority of those using the framework to behave as if a wheel is a stone cylinder and nothing else. However, a wheel is a general concept, which could refer to a stone cylinder, to a bicycle wheel, or to a wheel in a modern vehicle. Sometimes, even in a more modern vehicle, using a stone cylinder might be reasonable; often, it is disastrous.

3. Discussion

We have given a more detailed exposition of the probability expressions behind the faulty assumption that sensitivity and specificity do not depend on patient covariates, which has been discussed in Harrell et al. (2001); Moons et al. (1997); Guggenmoos-Holzmänn and van Houwelingen (2000). We have also discussed the clinical consequences associated with the assumption. In particular, we have shown that the assumption lures well-intentioned health-care providers into mixing and matching pre-test probabilities that depend on patient covariates with sensitivities and specificities that do not. By writing the probability calculations explicitly, we hope that we have provided more clarity on how the error is made and how it might be avoided.

It is easy to say how one should fix the error, but doing so would require work. One must conduct new, redesigned studies that estimate sensitivities and specificities that *do* depend on patient covariates; i.e., studies should estimate $p(\text{Test} = 1|Dz = 1)$ instead of $p(\text{Test} = 1|Dz = 1, x)$. Then, one could correctly compute (2). Harrell et al. (2001) describes (2) as taking three left turns to make a right. Even better, one could model each post-test probability, $p(Dz = 1|\text{test}, x)$, directly, without resorting to the use of Bayes rule. One would then have a risk score for the pre-test probability, $p(Dz = 1|x)$, and an additional risk score for the post-test probability, $p(Dz = 1|\text{test}, x)$. Evaluating the impact of a test would be tantamount to considering the change in post-test probability that would occur when moving from $p(Dz = 1|x)$ to $p(Dz = 1|\text{test}, x)$. In other words, if one has two tests, test_1 and test_2 , and one is not sure which one to order, one could

compute $p(Dz = 1|x, test_1)$ and $p(Dz = 1|x, test_2)$ and compare them both to $p(Dz = 1|x)$. The test that can change the post-test probability more in the desired direction should be chosen. This would be considerably more straightforward, more accurate, and less error prone than current practices.

4. Acknowledgments

This research, which is the sole responsibility of the authors and not the National Institutes of Health (NIH), was supported by the National Institute of General Medical Sciences (NIGMS) under T32GM007356.

References

- Goodacre, S., Sampson, F., Thomas, S., van Beek, E., Sutton, A., 2005. Systematic review and meta-analysis of the diagnostic accuracy of ultrasonography for deep vein thrombosis. *BMC medical imaging* 5, 1–13.
- Guggenmoos-Holzmamm, I., van Houwelingen, H.C., 2000. The (in) validity of sensitivity and specificity. *Statistics in medicine* 19, 1783–1792.
- Harrell, F.E., et al., 2001. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. volume 608. Springer.
- Hlatky, M.A., Pryor, D.B., Harrell, F.E., Califf, R.M., Mark, D.B., Rosati, R.A., 1984. Factors affecting sensitivity and specificity of exercise electrocardiography: multivariable analysis. *The American journal of medicine* 77, 64–71.
- Ledley, R.S., Lusted, L.B., 1959. Reasoning foundations of medical diagnosis: symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* 130, 9–21.
- Moons, K.G., van Es, G.A., Deckers, J.W., Habbema, D.J., Grobbee, D.E., 1997. Limitations of sensitivity, specificity, likelihood ratio, and bayes’ theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 8, 12–17.
- Moons, K.G., Harrell, F.E., 2003. Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. *Academic radiology* 10, 670–672.

- Parikh, R., Mathai, A., Parikh, S., Sekhar, G.C., Thomas, R., 2008. Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology* 56, 45.
- Prince-Guerra, J.L., Almendares, O., Nolen, L.D., Gunn, J.K., Dale, A.P., Buono, S.A., Deutsch-Feldman, M., Suppiah, S., Hao, L., Zeng, Y., et al., 2021. Evaluation of abbot binaxnow rapid antigen test for sars-cov-2 infection at two community-based testing sites—pima county, arizona, november 3–17, 2020. *Morbidity and Mortality Weekly Report* 70, 100.
- Trevethan, R., 2017. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in public health* 5, 307.
- Wells, P.S., Ginsberg, J.S., Anderson, D.R., Kearon, C., Gent, M., Turpie, A.G., Bormanis, J., Weitz, J., Chamberlain, M., Bowie, D., et al., 1998. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Annals of internal medicine* 129, 997–1005.