The Importance of Conditional Probability in Diagnostic Reasoning and Clinical Decision Making: A Primer for the Eye care Practitioner

PG Sanfilippo PhD,^{1,2} AW Hewitt PhD FRANZCO,^{1,2,3} DA Mackey MD FRANZCO^{1,2,3}

1. Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, Australia.
2. Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital.
3. School of Medicine, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia.

Manuscript Category: Review Article

Running Title: Diagnostic accuracy

Word Count: 4500

Corresponding Author

Dr Paul Sanfilippo

Centre for Eye Research Australia

32 Gisborne St, East Melbourne.

E-MAIL: prseye@gmail.com

Declaration of competing/conflicts of interest

No conflicts of interest were related to this work.

Financial Support

The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian Government. PGS is supported by an National Health and Medical Research Council (NHMRC) Early Career Fellowship.

Abstract

Purpose: To outline and detail the importance of conditional probability in clinical decision making and discuss the various diagnostic measures eye care practitioners should be aware of in order to improve the scope of their clinical practice.

Methods: We conducted a review of the importance of conditional probability in diagnostic testing for the eye care practitioner.

Results: Eye care practitioners use diagnostic tests on a daily basis to assist in clinical decision making and optimizing patient care and management. These tests provide probabilistic information that can enable the clinician to increase (or decrease) their level of certainty about the presence of a particular condition. While an understanding of the characteristics of diagnostic tests are essential to facilitate proper interpretation of test results and disease risk, many practitioners either confuse or misinterpret these measures.

Conclusions: In the interests of their patients, practitioners should be aware of the basic concepts associated with diagnostic testing and the simple mathematical rule that underpins them. Importantly, the practitioner needs to recognize that the prevalence of a disease in the population greatly determines the clinical value of a diagnostic test.

Introduction

Let us consider a hypothetical example in which a clever general practitioner (GP) identified a pharmacological biomarker for optic nerve disease. She subsequently developed a quick and inexpensive screening test for non-arteritic anterior ischemic optic neuropathy (NAION), motivated by her father's bilaterally poor visual outcome from the condition. The GP plans to administer the test to all patients over the age of 50 years as she does not wish anyone else to suffer the same fate. The blood test results are returned within minutes and are highly (90%) accurate. A random 60-year-old patient has the test and is informed the result is positive; should they be concerned? What is the probability the patient actually has NAION? Is there other information you need to know?

Intuitively, many of us might say that the chance is 90%, after all, the test is 90% accurate. Counter-intuitively, however, assuming an annual incidence rate for NAION in the general population of 10 per 100,000 individuals (0.01%),¹ the probability the patient actually has the disease (in a given year) is 0.09%! Thus, receiving a positive test raised the probability of the patient having NAION by 0.08%, an 8-fold relative increase, but not much in absolute terms. Such difficulties in probabilistic reasoning are common, for example, in one study investigating the accuracy of interpreting screening test information, over 50% of medical practitioners were incorrect in their responses.² In this case our intuition leads us to downplay or ignore the prior probability of the disease in favor of new evidence (ie, a positive test result) and create a cognitive bias known as base-rate neglect.³ Because we do not have all of the relevant information we can mistakenly jump to ill-conceived and often erroneous conclusions. We will deconstruct this problem later, but let us first review (in simple terms), a powerful mathematical rule that forms the foundation for the basic concepts associated with diagnostic testing.

Bayes theorem and why it's important to the practitioner

Proficiency in clinical reasoning and diagnostic test evaluation requires an understanding of the application of conditional probabilities in Bayes' theorem, a seminal language of chance ascribed to the Reverend Thomas Bayes (1701-1761).⁴ Conditional reasoning is used extensively in evidence-based medicine and the conditional probability simply reflects the probability of an event 'conditional upon' another event, or alternatively the probability of an event 'given that' another event has occurred. Thus, $P(B|A)$ represents the probability of B conditional upon A. Bayes' theorem is a mathematical rule that explains how one should change existing beliefs in light of new evidence and is the only correct way to associate conditional probabilities with their inverses. Thus, using Bayes' theorem it is possible to determine $P(B|A)$ from $P(A|B)$ if we also have information about the base rate or prior probability of events.

When a practitioner performs a diagnostic test, their primary motivation is to obtain information that assists in making a decision about the likelihood and diagnosis of a disease. It turns out, however, that the results of a diagnostic test represent the test probabilities of disease, not the real probabilities. Test probabilities of disease are affected by disease prevalence and test accuracy (the number of true positives, false positives, etc). The test provides $P(\text{test}|\text{disease})$ but what the clinician really wishes to know is $P(\text{disease}|\text{test})$, ie, what is the probability the patient has the disease given a positive test result. Bayes' theorem converts test probabilities into real probabilities of disease via Bayes' rule:

In general terms:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Applied in diagnostic reasoning:

$$P(D^+|T^+) = \frac{P(T^+|D^+)P(D^+)}{P(T^+)} \quad (1)$$

Where D^+ is Disease positive and T^+ is Test positive.

Thus, the probability that a patient has the disease given a positive test result [$P(D^+|T^+)$] is equal to the probability of a positive test result in the presence of disease [$P(T^+|D^+)$] multiplied by the probability of the disease in the population [$P(D^+)$, ie, the prevalence] divided by the overall probability of a positive test result [$P(T^+)$]. In essence, the probability of disease in light of a positive test result depends on the chance of a true positive result divided by the chance of any positive result. Figure 1 provides a graphical conceptualization of the relationship between conditional probabilities in diagnostic reasoning (five important diagnostic measures, which we will discuss in detail, are also shown).

While conditional probabilities underpin Bayes' theorem it is apparent from this figure that when test/disease frequencies and diagnostic measures are framed in probabilistic terms, their interpretation appears intimidating. This highlights a fundamental problem in clinical medicine whereby many practitioners not only fail to understand the difference in conditional probabilities as applied to diagnostic tests but when presented with the information do not know how to compute the relevant measure. Alternatively,

research has shown that representing statistical information in terms of natural frequencies rather than probabilities improves competency in Bayesian inference tasks.⁵ Humans are intuitively more adept in dealing with natural frequencies as raw numbers have existed considerably longer than probability theory and unlike conditional probabilities they are not normalized with respect to the base rates of the event in question (ie, while the marginal probabilities in Figure 1 add to 100%, the four conditional probabilities do not). Figure 2 shows how the same information may be presented in terms of natural frequencies rather than probabilities. Ignoring the summary columns and rows, this is a simple 2x2 contingency table showing the counts of individuals with disease who have tested positive (true positive, TP), with disease who have tested negative (false negative, FN), without disease who have tested positive (false positive, FP) and without disease who have tested negative (true negative, TN). Knowledge of these four count values enables the practitioner to compute any of the suite of diagnostic measures essential to the evaluation of test efficacy and/or disease risk.

Basic concepts in diagnostic testing

Diagnostic tests are both ubiquitous and necessary in clinical medicine to identify the presence or absence of disease and develop an appropriate treatment plan. When a new test ('index test') is developed, a measure of its diagnostic accuracy is determined by evaluating its agreement with an existing diagnostic test ('gold standard' or 'reference test'), considered the best test (or combination of tests) currently available in detecting the disease of interest. In Figures 1 and 2, 'Patient's True State' reflects the diagnosis determined by the reference test and 'Test Result' the index test that is being compared. For example, meta-analysis of multiple diagnostic accuracy studies of fundus fluorescein

angiography (FFA) might reveal the procedure can tell us with great certainty whether or not a patient has neovascular age-related macular degeneration (AMD). With the advent of high-resolution optical coherence tomography (OCT) we may wish to evaluate its utility in detecting the disease. In this case, the 'Patients True State' would reflect the result of the FFA and the 'Test Result' being compared would be that of the OCT.

Sensitivity and specificity

The diagnostic accuracy of a test is characterized by two measures of conditional probability: Sensitivity and specificity (Figures 1 and 2). Of the total patient sample, the sensitivity of a test concerns only those individuals who have disease and specificity only those who do not. Sensitivity measures the proportion of people with disease that are correctly classified by the test as having disease. In conditional probability terms it is the proportion of people who have tested positive given they have the disease. In contrast, specificity measures the proportion of people without disease that are correctly identified as such by the test. Probabilistically, it is the proportion of people who have tested negative given they are disease free. Alternatively, let us consider their natural frequency representations. In this interpretation, sensitivity is simply the number of TPs divided by the total number of people with disease (TP + FN; Figure 2). Here, specificity is the number of TNs divided by the total number of people without disease (FP + TN). A very simple 2x2 contingency table will help to illustrate these concepts (Figure 3). In the evaluation of OCT as a diagnostic tool in glaucoma, 100 people have both retinal nerve fiber layer (RNFL) thickness measurements performed (index test) and a comprehensive ophthalmic and perimetric assessment by a glaucoma specialist (reference test). For the sake of simplicity, the resulting RNFL thickness measurement is converted to a binary variable indicating the presence or absence of glaucoma. We are provided with the four

principal count values. In this example, therefore, the sensitivity of RNFL thickness determination to primary open-angle glaucoma (POAG) is $4/5=80\%$. Similarly, the specificity of RNFL thickness determination to POAG is $76/95=80\%$.

Likelihood ratios

Likelihood ratios (LR) represent alternative statistics for summarizing diagnostic accuracy and evaluating the likelihood of disease. In general, LRs are a ratio of the likelihood of a specific test result (positive or negative) in patients *with* the disease, divided by the likelihood of the *same* test result (positive or negative) in patients *without* the disease. As there are two possible test results (for a dichotomous test), there are also two LRs (positive and negative):

$$LR+ = \frac{TP \text{ rate}}{FP \text{ rate}} = \frac{\text{sensitivity}}{1 - \text{specificity}} = \frac{TP}{TP + FN} \div \frac{FP}{FP + TN}$$

$$LR- = \frac{FN \text{ rate}}{TN \text{ rate}} = \frac{1 - \text{sensitivity}}{\text{specificity}} = \frac{FN}{TP + FN} \div \frac{TN}{FP + TN}$$

The positive LR (LR+) compares the probability of a positive test result given the *presence* of disease to the probability of a positive test result given the *absence* of disease. The negative LR (LR-) compares the probability of a negative test result given the *presence* of disease to the probability of a negative test result given the *absence* of disease. It is important to note that the LR is essentially an odds ratio and thus expresses a change in odds, although conversion to probabilities is a relatively simple undertaking.⁶

In our earlier example (Figure 3), if a patient had a positive test result (thin RNFL indicating glaucoma), the LR+ would be calculated as 4. On the other hand, if the result

was negative (normal RNFL thickness), the LR- would be computed as 0.25. In the former case, the likelihood of the patient having POAG has increased by approximately 4-fold and in the latter, decreased by the same. Aside from LR+ being natural to interpret (larger LR+ = disease more likely, smaller LR- = disease less likely), they have other useful properties. Because they are a ratio of sensitivity and specificity, they are not population or setting dependent (as sensitivity and specificity are) and are generalizable. In addition, LR+ can deal with ordinal or continuous test data (not just normal/abnormal) and are able to be utilized directly at the individual patient level. Importantly, they don't depend on the prevalence of disease, as predictive values do.

Positive and negative predictive values

The appealing characteristics of the LR in diagnostic test evaluation have largely gone underutilized for patient care. While the LR itself is intuitive, the ability to easily determine post-test disease probabilities is not and requires the use of nomograms or additional calculations (conversion of pre-test probability to pre-test odds, computation of post-test odds, conversion of post-test odds to post-test probability). This has been suggested as a cause for their limited use in clinical practice.⁷ In this sense, the use of predictive values is most straightforward as the disease post-test probability is equivalent to the predictive value itself.

The predictive value of a test enables the practitioner to answer the question that is of most relevance to both themselves and the patient: What is the probability of having (not having) a particular disease given a positive (negative) test result? The positive predictive value (PPV) represents the probability of having the disease given a positive test result $P(D^+ | T^+)$, and the negative predictive value (NPV), the probability of *not* having the disease given a negative test result $P(D^- | T^-)$. Predictive values in their conditional probability and

natural frequency representations are shown in Figures 1 and 2, respectively. Under the latter interpretation, the PPV is simply the number of TPs divided by the total number of people who tested positive (TP + FP) and the NPV is the number of TNs divided by the total number of people who tested negative (FN + TN; Figure 2). Alternative, yet equivalent, representations of the PPV and NPV (when the sensitivity, specificity and disease prevalence are known) are based on application of Baye's theorem (1):

$$PPV = \frac{\text{sensitivity} * \text{prevalence}}{\text{sensitivity} * \text{prevalence} + (1 - \text{specificity}) * (1 - \text{prevalence})}$$

$$NPV = \frac{\text{specificity} * (1 - \text{prevalence})}{(1 - \text{sensitivity}) * \text{prevalence} + \text{specificity} * (1 - \text{prevalence})}$$

In our example (Figure 3), the PPV of OCT-determined RNFL thickness for POAG would be calculated as 17% (4/23) and the NPV, 99% (76/77). Since a very high NPV and a negative test result virtually rules out disease, we can feel confident that a normal RNFL thickness measurement rejects the diagnosis of glaucoma. The opposite cannot be said, however, if an abnormal RNFL thickness result is returned. In this case, if a patient has a thin RNFL thickness on OCT imaging the probability they actually have glaucoma is still only 17%.

Wider application of conditional probability in diagnostic testing

A perfect test would correctly classify all with disease (TP) and all without (TN) and in this case the diagnostic accuracy would be 100%. However, tests are prone to errors and it is not unusual that they miss diagnoses (FN) and/or incorrectly indicate the presence of disease (FP). Normally, a desirable characteristic of a test is that it minimizes the

number of FPs and FNs, thus maximizing the sensitivity and specificity. But this is often not possible and in practice the sensitivity and specificity of a test are intertwined such that a trade-off occurs between the two; increasing one decreases the other. Indeed, there may even be circumstances where it is preferable to use a test (if possible) that optimizes one measure over the other, depending on the consequences of having FPs or FNs. For example, if on balance, the seriousness of FNs outweigh FPs and we wish to avoid the former, then a test with high sensitivity would be of most value (Figure 2; minimize FNs, maximize TPs). This would be the case where the penalty for missing a diagnosis is important and could result in vision loss or death (eg, central retinal artery occlusion, ocular tumors, etc). If, on the other hand, a FP error is of more significance to the patient, as might be the case if it leads to an unnecessary follow-up procedure that is expensive and/or invasive, then it would be sensible to optimize specificity (minimize FPs, maximize TNs). It is worth noting that our discussion has been based on a binary outcome; the presence or absence of disease. However, the same principles may be extended to accommodate a continuous outcome where a series of sensitivity and specificity paired values are computed for multiple thresholds of the outcome variable (eg, sensitivities and specificities calculated at different levels of intraocular pressure). In this way a receiver operating characteristic (ROC) curve may be constructed which provides an overall measure (across all possible thresholds) of the diagnostic performance of a test.

These concepts may be further developed to provide two basic rules of thumb in helping the practitioner apply sensitivity and specificity information in clinical decision making. At first glance they may seem non-intuitive but considered inspection of an example 2x2 contingency table (eg, Figure 3) will help to consolidate the general idea in the readers mind. By definition a highly sensitive test will produce few FNs and so it follows that a negative result from a sensitive test is most likely to be a TN. In other words, a very

sensitive test (even if not very specific) is most useful when negative and provides a degree of confidence that the patient does *not* have the disease. Conversely, a highly specific test will generate few FPs and thus a positive result is more likely to be true. In this case, a very specific test (even if not particularly sensitive) will be most helpful when positive, providing confidence that the patient *does* have the disease. David Sackett, instrumental in the evidence-based medicine movement, coined two mnemonics to help remind the practitioner of these properties⁸:

1. **‘SnNout’** - **S**ensitive test, **N**egative result, rules **out**.

That is, a sensitive test (when negative) helps rule out disease.

2. **‘SpPin’** - **S**pecific test, **P**ositive result, rules **in**.

Or, a specific test (when positive) helps rule in disease.

For example, SnNout can be applied when evaluating the loss of spontaneous venous pulsation (SPV) as a sign of raised intracranial pressure (ICP). SPV is present in about 90% of the normal population. Because it is absent in the remaining 10%, lack of SPV does not confirm ICP is raised. However, no SPV (ie, a positive test result) is very sensitive ($\sim 90\%$)⁹ for raised ICP, therefore its presence (ie, a negative test result) helps to rule out the diagnosis of papilledema. In contrast, we can apply the principle of SpPin to the results of a recent systematic review of the diagnostic accuracy of the Amsler grid in the screening of patients with AMD.¹⁰ In a meta-analysis of 12 studies, Faes et. al calculated pooled sensitivities and specificities of 78% and 97%, respectively. One way of interpreting these data is that while the sensitivity is good, 22% of people with AMD may still not notice any distortion (ie, a positive test result) on an Amsler grid. The test is

therefore not as useful in ruling out the disease. With such a high specificity though, the presence of distortion does help to rule in the diagnosis of AMD.

While these principles provide convenient clinical guidelines, various authors have cautioned against their use as an ironclad interpretation of diagnostic test results.^{11, 12} A critical appraisal of the evidence as well as the methodological quality of the test study are also necessary in evaluating the claimed efficacy of a new test. In addition, the power to rule a disease in or out is diminished when highly specific tests are not sufficiently sensitive, or highly sensitive tests are not sufficiently specific.¹¹ For this reason, the calculation and reporting of likelihood ratios have been advocated. The likelihood ratio depends on both sensitivity and specificity and hence takes into account the trade-off between them.

But how does the LR inform our clinical decision making? As we alluded to earlier, when a practitioner performs a diagnostic test with a particular disorder in mind they are hoping the result will modify their initial assessment of the pre-test probability of the disorder in some way. This 'shift in suspicion' can lead them to be more certain about the post-test probability of disease (positive) or less certain (negative). In other words, LRs tell us how much we should shift our suspicion for a particular test result. LRs may range from zero to infinity. When the LR is equal to 1, the test lacks diagnostic value as the post-test probability of disease is the same as the pre-test probability. As the LR moves further away from 1 and becomes more positive or more negative, the strength of the evidence for the presence or absence of disease, respectively, increases. Thus, LR+ corresponds to the concept of ruling in disease and LR- to the concept of ruling out disease. To avoid confusion, it is worthwhile noting that the LR always indicates the likelihood of disease; the positive and negative label simply refers to the test result. As a general guideline, LRs

above 2 or below 0.5 provides minimal evidence for a test that may help rule in or rule out a diagnosis.

Conditional probabilities and the base rate fallacy

If we now return to Figure 3, why is it that a test with good diagnostic accuracy (80%) performs so poorly in correctly identifying patients with disease? As the reader has probably surmised, the answer has to do with the base rate or prevalence of the disease the test is being applied to.

Base rate neglect (“base rate fallacy”, “inverse fallacy”, or the “prosecutor’s fallacy” in legal reasoning) occurs when we are trying to make a conditional probability judgement and incorrectly assume that $P(A|B) = P(B|A)$. Let us imagine that $P(\text{Golf} | \text{Eye care Practitioner})$ represents the probability that one plays golf given they are an eye care practitioner. This example clearly illustrates the error in assuming equiprobability in transposing conditional events. The probability that one plays golf given they are an eye care practitioner, ie, $P(\text{Golf} | \text{Eye care Practitioner})$, is probably close to 100%. However, the probability of being an eye care practitioner given that you play golf, $P(\text{Eye care Practitioner} | \text{Golf})$, is substantially lower. On a more serious note, consider the probability of being male, given that you are diagnosed with Leber hereditary optic neuropathy (LHON; $P(\text{Male} | \text{LHON})$). As males are much more commonly affected, this may be as high as 90%, yet the conditional inverse, ie, $P(\text{LHON} | \text{Male})$ remains low. The salient point here is to be aware of the ‘base rate’ or prior probability (ie, prevalence, incidence, etc) of events. The proportion of Eye care Practitioners among the general public is quite low but the proportion of golfers is not. The chances of any individual

developing LHON are similarly quite low, but about half the population are male. Thus, if the prior probability of an event is very low, then the revised ('posterior') probability even after new evidence has come to light will still be low (here we refer to 'new evidence' as the imposed condition of playing golf or being male). It is important to be aware of this distinction as the base rate fallacy and conditional probabilities play an integral role in diagnostic reasoning and the evaluation of clinical evidence. Clinical decision making may be challenging because the practitioner is often required to assess the evidentiary weight for an event that is contingent on the occurrence of another (for example, assessing the risk of retinal detachment in a patient needing cataract surgery given they are highly myopic).

Base rates determine test value

Let us now return to our opening problem. So what is the probability the patient actually has NAION? In other words, what is the PPV of the hypothetical blood test in detecting this relatively rare neuropathy? As explained in the previous section, this is a very simple calculation if provided with natural frequencies. All we need are the number of TPs and FPs and we can calculate the PPV. However, this may not be our experiment and we may not have easy access to these data. Instead we are provided with test sensitivities and specificities (or their complements, the FN and FP rates, respectively). The third and final piece of information needed to evaluate test utility is the base rate or prior probability (ie, prevalence, incidence, etc) of the disease. We can build on the calculations performed in our simple example (Figure 3) by essentially working backwards to compute natural frequencies (Figure 4) from the supplied information:

1. We are told the accuracy of the blood test is 90%. In the absence of additional information it is fair to assume that both the sensitivity and specificity are therefore the same (90%).

2. With an annual incidence rate for NAION of 0.01% we can convert this to a natural frequency of 10 per 100,000 individuals (in this case choosing a total of 100,000 ensures we maintain whole numbers when we compute the count values).

3. The total number of people with disease (in a given year) is therefore 10 and this corresponds to the marginal total D^+ (TP + FN).

4. Of those 10 people, 9 are correctly identified as having disease (sensitivity = 90%). One person is incorrectly classified as being disease-free. The number of TPs and FNs is thus 9 and 1, respectively.

5. The total number of people without disease (in a given year) is 99,990 (100,000-10) and this corresponds to the marginal total D^- (FP + TN).

6. Of those 99,990 people, 89,991 are correctly identified as being disease free (specificity = 90%). This represents the number of TNs. The complement is 9,999 (99,990-89,991) and indicates the number of individuals who are incorrectly classified as having disease, or the number of FP test results.

As we now know the four count values we can easily calculate the PPV and NPV. The revised (posterior) probability of having NAION conditional on a positive test (PPV)

has increased from 0.01% to 0.09%, but the test remains ineffective in correctly identifying individuals with disease. Because of the very low incidence of NAION, there are many more FPs than TPs and for such rare diseases, a large proportion of those with positive screening tests will necessarily be found not to have disease upon further diagnostic testing. In short, if a patient has a positive test result the probability they actually have the disease is very low. At the same time, the low incidence of NAION means that negative test result is more accurate. Here the NPV is almost 100%, thereby allowing the practitioner great confidence in reassuring their patient with a negative result that they are disease free. The salient point from this example is that even a highly ‘accurate’ test can have a low PPV if the disease is rare. In diagnostic medicine, the base rate fallacy occurs when the prior probability of disease in the population is so low that the vast majority of the people undertaking the test are healthy. Consequently, even with an accurate test, many of the individuals who test positive will be disease free, confounding the correct interpretation of disease risk.

To conclude, let us consider how the predictive value is affected when the disease is more common. Imagine the year is 2216 and the planet’s demography has skewed strikingly towards the aged. Humans are now routinely living until the age of 120 years and the proportion of individuals aged over 50 years is greater than their younger counterparts. Unfortunately, we have still not found a cure for POAG and the prevalence of the disease has increased to about 50% (from $\sim 4\%$ today). A new diagnostic test has been devised that is similarly 90% accurate. What is the predictive value of the test? It is left to the reader to follow the workings described previously but we can see from Figure 5 that a test with equivalent accuracy performed in the setting of a more common disease is of

substantially greater utility in correctly classifying those who have tested positive with disease (PPV has increased to 90%).

While the focus of this paper and the discussion of conditional probability has been in the context of medical diagnostic testing, the base rate fallacy is widely applicable in many areas of probabilistic judgement. For example, the ‘prosecutor’s fallacy’ occurs in legal reasoning whereby conditional probabilities are confused and/or misappropriated by the prosecution to argue for the guilt of a defendant during a criminal trial. The majority of Australians drive, yet only a very small proportion of individuals drive while drunk. If an individual receives a positive random breath test, what is the probability they are actually drunk? In an effort to enhance aviation security, criminal intent prescreening has been implemented by the Transportation Security Administration (TSA) in the USA. Given the prevalence of terrorists is undoubtedly low, if the TSA’s predictive algorithm identifies a random individual as being of concern, what is the probability they are in fact an airplane terrorist?

Bayesian reasoning plays a role in daily life and most of us are unaware we are incorporating Bayes theorem into our everyday decision-making. We are constantly re-evaluating what we know based on previous experience and/or new information. It is important when making a judgement about an event that the base rate of the event be considered, and this is especially prudent for the clinician. Recognizing the difference in conditional probabilities and their interpretation is essential in providing patients with the correct information and clinical advice for their best care.

Legend

Figure 1: Graphical illustration of relationships between conditional probabilities in diagnostic reasoning.

Figure 2: Diagnostic summary data represented as natural frequencies.

Figure 3: Diagnostic measures of optical coherence tomography (OCT) retinal nerve fiber layer (RNFL) thickness as a screening tool in glaucoma (n=100).

Figure 4: Diagnostic measures of a hypothetical blood test as a screening tool in non-arteritic anterior ischemic optic neuropathy (NAION; n=100,000, prevalence =0.1%).

Figure 5: Diagnostic measures of a hypothetical new test as a screening tool in glaucoma (n=100,000, prevalence =50%).

References

1. Hattenhauer MG, Leavitt JA, Hodge DO, Grill R, Gray DT. Incidence of nonarteritic anterior ischemic optic neuropathy. *Am J Ophthalmol* 1997;123(1):103-7.
2. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ* 2006;333(7562):284.
3. Bar-Hillel M. The base-rate fallacy in probability judgments. *Acta Psychol (Amst)*, 1980.
4. Bayes T, Price R. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. *Philosophical Transactions of the Royal Society of London* 1763;53370-418.
5. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol Rev* 1995;102684-704.
6. Spitalnic S. Risk Assessment II: Odds Ratio. *Hosp Physician*, 2006;23.
7. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*, 2002;646-649.
8. Sackett D. Clinical Epidemiology: A Basic Science for Clinical Medicine. Boston: Little Brown, 1991.
9. Wong SH, White RP. The clinical validity of the spontaneous retinal venous pulsation. *J Neuroophthalmol* 2013;33(1):17-20.
10. Faes L, Bodmer NS, Bachmann LM, Thiel MA, Schmid MK. Diagnostic accuracy of the Amsler grid and the preferential hyperacuity perimetry in the screening of patients with age-related macular degeneration: systematic review and meta-analysis. *Eye (Lond)* 2014;28(7):788-96.

11. Pewsner D, Battaglia M, Minder C, Marx A, Bucher HC, Egger M. Ruling a diagnosis in or out with 'SpPIN'; and 'SnNOut': a note of caution. *BMJ: British Medical Journal Publishing Group*, 2004:209-213.
12. BOYKO EJ. Ruling Out or Ruling in Disease with the Most Sensitive or Specific Diagnostic-Test - Short-Cut or Wrong Turn. *Med Decis Making*, 1994:175-179.