Concept article

When math legitimizes knowledge: a step by step approach to Bayes' rule in diagnostic reasoning

Yung Bruno de Mello Gonzaga[1]
André Demambre Bacchi[2]
Vitor Borin Pardo de Souza[3]

[1]Corresponding author. Instituto Nacional de Câncer (Rio de Janeiro). Rio de Janeiro, Brazil. Grupo Oncoclínicas (Rio de Janeiro). Rio de Janeiro, Brazil. ygonzaga@inca.gov.br
[2]Universidade Federal de Rondonópolis (Rondonópolis). Mato Grosso, Brazil.
[3]Universidade do Estado de São Paulo (São Paulo). São Paulo, Brazil.

ABSTRACT | INTRODUCTION: Many mistakes in clinical practice arise from confusing the probability of a positive test in those with the disease and the probability of having the disease in those who test positive. This misunderstanding leads to overestimating disease probability, diagnosing diseases in healthy individuals, ordering invasive diagnostic tests, and prescribing unnecessary treatments, resulting in unjustified adverse effect, psychological stress, and increased cost. Probabilistic reasoning is an essential skill to mitigate this confusion, and Bayes theorem is an important tool to accomplish this goal. **OBJECTIVE:** To present a step-by-step demonstration of Bayes' formula for positive and negative predictive values, fostering understanding and enabling its adoption in evidence-based medicine education and clinical practice as a supporting tool in the decision-making process. **METHODS:** In this article, we explain the difference between deductive and inductive thinking and how diagnostic reasoning is predominantly inductive, where evidence (the test result) is used to predict the cause (the presence of disease), a path that involves reverse probability, for which our reasoning is hazier. Through a clinical example involving the diagnosis of systemic lupus erythematosus, we use the Bayesian framework as a tool to help understand the difference between sensitivity/specificity (forward probability; deductive) and positive/negative predictive values (reverse probability: inductive). **CONCLUSIONS:** Excellent doctors are masters at applying Bayesian reasoning without using any formulas: they understand that the most important component of the diagnostic process is the reasoning that originates it and the resulting clinical decision depends on interpreting results considering their interaction with the context, not in isolation. Bad clinical reasoning results in bad clinical decisions, despite how accurate the diagnostic test: garbage in, garbage out. We hope our step-by-step approach to Bayes' rule can help demystify this powerful statistical tool and strengthen the idea that the value of a diagnostic test is directly proportional to the quality of clinical reasoning that led to its request.

KEYWORDS: Probability. Clinical Decision-Making. Diagnostic Errors.

Submitted Jan. 1st, 2024, Accepted Oct. 18th, 2024, Published Nov. 19th, 2024
J. Évid-Based Healthc., Salvador, 2024;6:e5903
http://dx.doi.org/10.17267/2675-021Xevidence.2024.e5903 | ISSN: 2675-021X
Assigned editors: João de Deus Barreto Segundo, Luís Cláudio Correia

How to cite this article: Gonzaga YBM, Bacchi AD, Souza VBP. When math legitimizes knowledge: a step by step approach to Bayes' rule in diagnostic reasoning. J Évid-Based Healthc. 2024;6:e5903. http://dx.doi.org/10.17267/2675-021Xevidence.2024.e5903

CC BY

1. Introduction

Many mistakes in clinical practice arise from confusing the probability of a positive test in those with the disease and the probability of having the disease in those who test positive. This misunderstanding leads to overestimating disease probability[1], diagnosing diseases in healthy individuals[2], ordering invasive diagnostic tests[3,4] and prescribing unnecessary treatments, resulting in unjustified adverse effects[5,6], psychological stress[7] and increasead costs.[8] Probabilistic reasoning is an essential skill for doctors. To apply it, besides specific medical knowledge, an understanding of some basic mathematical and statistical concepts is required.[9,10] One of these concepts is Bayes' theorem, developed by the mathematician Pierre-Simon Laplace, based on Thomas Bayes' work[11], an 18th century English pastor.

$$Bayes\ theorem: \qquad P(A|B) = \frac{P(A)\ x\ P(B|A)}{P(B)}$$

Bayesian reasoning can be summarized as:

prior knowledge (or belief) + new evidence → updated knowledge (or belief).

It provides an objective way to combine the result of a diagnostic test (new evidence) with initial clinical suspicion (prior knowledge), updating knowledge and decreasing diagnostic uncertainty.

The theorem has been adapted for diagnostics, and is referred to as Bayes' rule for positive and negative predictive values (PPV and NPV):[12,13]

$$PPV = \frac{(prevalence\ x\ sensitivity)}{[(prevalence\ x\ sensitivity)] + [(1 - prevalence)\ x\ (1 - specificity)]}$$

$$NPV = \frac{[(1 - prevalence)\ x\ specificity]}{[(1 - prevalence)\ x\ specificity] + [prevalence\ x\ (1 - sensitivity)]}$$

The seemingly complex formula's visual impression, coupled with the absence of an explanation of how it was derived, may lead to resistance to adopting it.[14] Therefore, we present a step-by-step demonstration integrated with clinical reasoning, which can serve as a didactic tool for those who teach evidence-based medicine and those who want to apply it in clinical practice.

2. Probability theory

Probability theory provides formal and mathematical language to deal with events that involve uncertainty, such as diagnosis. What truly interests us is a dichotomous question: whether the individual has the disease being investigated or not. The problem arises when we also interpret test results as dichotomous: positive test = disease; negative test = absence of disease. Test results rarely settle the issue in such a definitive manner. What they do is reduce uncertainty, relying on probability.

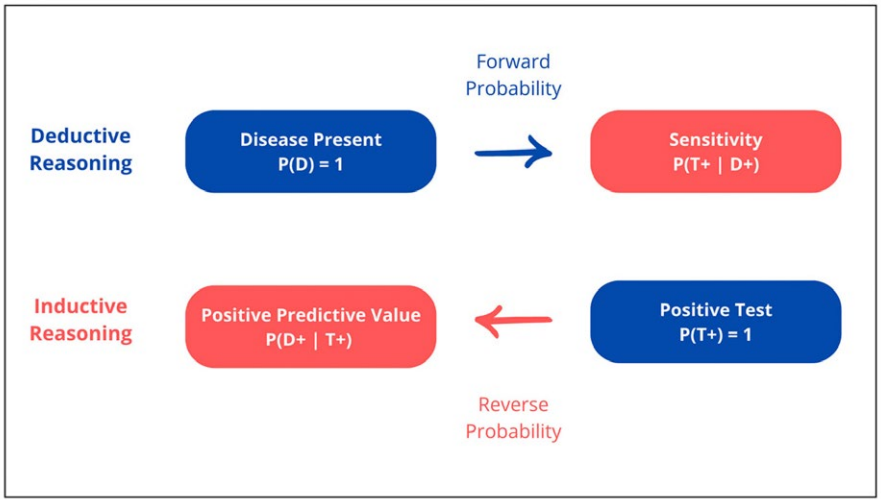
But what does a 70% probability of having a disease mean? Or a 20% risk of death? Nobody is 70% sick or 20% dead. Probability continuously quantifies categorical events, which is why it is so difficult to understand. Probability doesn't exist as a tangible entity; it is an abstraction.

3. Diagnostic reasoning: inductive vs. deductive

The probability of a positive result in someone who has the disease is the test sensitivity.[15] This probability progresses from cause (disease) to effect (test result), a natural way of thinking, called deductive. However, in clinical practice, we do not know if the patient has the disease (if we did, we would not need a test). We start from uncertainty: a clinical suspicion motivates the solicitation of the test and, once we have the result, we should think in the opposite direction: the probability that the positive test result (the effect) truly indicates the presence of disease (the cause): this probability is the PPV. It is an unnatural way of thinking, one for which our reasoning is hazier: a problem of reverse probability or inductive thinking. The same rationale applies to specificity (the probability of a negative test given the absence of disease) and NPV (the probability of not having the disease given a negative test).[15]

Diagnostic clinical reasoning is, therefore, predominantly inductive (Figure 1), and Bayes' rule helps us deal with the complex issue of reverse probability and estimate what truly matters: whether the patient has the disease or not, based on the interaction between the observed evidence (diagnostic test result) and our prior knowledge (initial clinical suspicion). Let us consider a clinical example.

Figure 1. Deductive and inductive thinking in the diagnostic process



Legend: Individuals previously diagnosed with the disease through a gold standard test are subjected to a new test to determine its sensitivity. The test's sensitivity is calculated by measuring the proportion of positive results among these diseased individuals. This process involves forward probability and incorporates aspects of deductive reasoning. However, in clinical practice, clinicians often face the challenge of reasoning from the opposite direction. They need to estimate whether a patient with an unknown disease status is truly diseased, based on a positive test result. This type of reasoning involves reverse probability. It primarily employs inductive methods and Bayesian reasoning, enabling clinicians to update hypotheses (clinical suspicion) based on the available evidence (test result).
Source: the authors (2024).

4. Applying Bayesian reasoning to the diagnosis of Systemic Lupus Erythematosus (SLE)

SLE is a severe autoimmune disease characterized by autoantibodies against nuclear antigens, immune complex deposition, and chronic inflammation of target organs. Despite advances in treatment, this disease continues to have high morbidity and mortality rates.[16] One of the main tests used in the diagnosis of SLE is the antinuclear antibody (ANA) test. It has a sensitivity and specificity of approximately 96% and 86%, respectively, at a titer of 1:160.[17] Despite being an accurate test for diagnosing a lethal and morbid disease, the use of the ANA test is not recommended for screening SLE in asymptomatic individuals[18] and Bayesian reasoning can help us understand why.

Assuming a prevalence of SLE of 50 cases per 100,000 individuals[17] or 0.05% and applying these in Bayes' formula, we have:

$$PPV = \frac{(0.05\ x\ 96)}{(0.05\ x\ 96) + (99.5\ x\ 14)}$$

$$PPV = 0.3\%$$

$$NPV = \frac{(0.95\ x\ 86)}{(0.95\ x\ 86) + (0.05\ x\ 4)}$$

$$NPV = 99.9\%$$

A PPV of 0.3% means the test has a 99.7% probability of being a false positive.

An NPV of 99.99% means the test has almost 100% probability of being a true negative. Therefore, in this context, both the positive and negative results represent virtual certainty of the absence of disease. In the case of the positive test, misinterpretation as true positive can lead to bad clinical decisions and the negative test, although correct, ruled out something we should already know. In summary, for an asymptomatic individual, the ANA test result ranges from zero benefits to potential harm and should not be ordered.

Now, suppose a young female patient comes to the office with a history of fatigue, arthralgia, and skin lesions. After a thorough history and physical examination, the clinician suspects that SLE is among the differential diagnoses and estimates a 30% probability for the disease. This value corresponds to uncertainty, which makes us uncomfortable with assuming the diagnosis to the point of starting a treatment, but also with rejecting it and not identifying a serious disease needing treatment. This uncertainty can be reduced by requesting an ANA test. Let us examine how the test would perform if performed by applying the values in Bayes' formulas:

$$PPV = \frac{(30\ x\ 96)}{(30\ x\ 96) + (70\ x\ 14)}$$

$$PPV = 74.6\%$$

$$NPV = \frac{(70\ x\ 86)}{(70\ x\ 86) + (30\ x\ 4)}$$

$$NPV = 98\%$$

A positive test result made the diagnosis more likely (PPV of 74,6%) and will motivate the rational request for additional tests, while a negative test result practically ruled out SLE (NPV of 98%) and will direct the approach towards finding alternative diagnoses.

Based on a clinical suspicion derived from a thorough medical history and targeted physical examination, the result of the same test proved to be extremely useful in the decision-making process.

5. Bayes' rule and predictive values

After this example, let's demonstrate the step-by-step development of Bayes' formulas for PPV and NPV.

5.1 Bayes' rule for PPV

We can conclude that a positive test is a true positive in two ways: through the deductive path or the inductive path.

Deductive path: For a positive test result to be a true positive, two things must occur: the patient must have the disease AND the test must be positive for this patient. We will express the probability of having the disease before any test result as P(D+), the pre-test probability (or prevalence) of the disease. We will express the probability of the test being positive given the presence of the disease as P(T+|D+), the sensitivity of the test. Since we are dealing with two independent conditions for a positive test result to qualify as a true positive, we multiply the probabilities:

$$P(D+)\ x\ P(T+|D+)$$

Inductive path: For a positive test result to be a true positive, two things must occur: the test must be positive AND the patient with a positive test must have the disease. We will express the probability of the test being positive (regardless of whether it is a true or false positive) as P(T+) and the probability of having the disease given the positive test as P(D+|T+), the PPV.

Since we are dealing with two independent conditions for a positive test result to qualify as a true positive, we multiply the probabilities:

$$P(T+)\ x\ P(D+|T+)$$

As both expressions are mathematically the same, we can represent them as equals:

$$P(T+)\ x\ P(D+|T+) = P(D+)\ x\ P(T+|D+)$$

As previously mentioned, one of the main misconceptions in the interpretation of diagnostic tests is the mix between the probability of a positive test in someone who has the disease, i.e., P(T+|D+) and the probability of having the disease in someone who has a positive test, i.e., P(D+|T). They are different, yet related, probabilities. We can understand this relationship by moving P(T+) to the other side of the equation isolating the component that is relevant for decision-making and obtaining the original structure of Bayes' theorem:

$$P(D+|T+) = \frac{P(D+)\ x\ P(T+|D+)}{P(T+)}$$

Replacing in the formula:

P(D+|T+) is the PPV;
P(D+) is the prevalence or pre-test probability;
P(T+|D+) is the sensitivity;
P(T+) is the probability of any positive test, that is, true positives OR false positives. Understanding this component: for a test to be a true positive, the disease must be present, and the test must be positive in the presence of the disease: P(D+) x P(T+|D+). On the other hand, for a false positive, the patient must not have the disease P(D-), and the test must be positive in the patient without the disease P(T+|D-). Hence, the probability of a positive test is the sum of the probabilities of true positives and false positives as follows:

$$P(T+) = [P(D+)\ x\ P(T+|D+)] + [P(D-)\ x\ P(T+|D-)]$$

where:

P(D-) is the probability of not having the disease, or: 1 - P(D+)
P(T+|D-) is the probability of a positive test given that there is no disease. This is the complement of specificity, that is: 1 – P(T-|D-).

Rewriting the theorem, substituting the notations of conditional probability with the terms used in the diagnostic process, we obtain Bayes rule for the PPV. So:

$$P(D+|T+) = \frac{P(D+)\ x\ P(T+|D+)}{[P(D+)\ x\ P(T+|D+)] + [(1-P(D+))\ x\ (1-P(T-|D-))]}$$

becomes:

$$PPV = \frac{(prevalence\ x\ sensitivity)}{[(prevalence\ x\ sensitivity)] + [(1-prevalence)\ x\ (1-specificity)]}$$

5.2 Bayes' rule for NPV

The development of the formula for the Negative Predictive Value (NPV) follows the same rationale. We can conclude that a negative test is a true negative in two ways: through the deductive path or the inductive path.

Deductive path: For a negative test result to be a true negative, two things must occur: the patient must not have the disease AND the test must return negative for this patient without the disease. We express the probability of not having the disease before any test result as P(D-) and the probability of the test being negative given the absence of the disease as P(T-|D-), the specificity of the test. Since we are dealing with two independent conditions for a negative test result to qualify as a true negative, we multiply the probabilities:

$$P(D-)\ x\ P(T-|D-)$$

Inductive path: We can obtain the same result by taking the reverse path, starting from the test. For a negative test result to be a true negative, two things must occur: the test must be negative AND the patient with negative test must not have the disease. We express the probability of the test being negative (regardless of whether it is a true or false negative) as P(T-), and the probability of not having the disease given the negative test as P(D-|T-), the NPV.

Since we are dealing with two independent conditions for a negative test result to qualify as true negative, we multiply the probabilities:

$$P(T-)\ x\ P(D-|T-)$$

As both expressions are mathematically the same, we can represent them as equals.

$$P(T-)\ x\ P(D-|T-) = P(D-)\ x\ P(T-|D-)$$

If we move P(T-) to the other side of the equation and isolate the component that is relevant for decision-making, we obtain the original structure of Bayes' theorem:

$$P(D-|T-) = \frac{P(D-)\ x\ P(T-|D-)}{P(T-)}$$

Replacing in the formula:

P(D-|T-) is the NPV;
P(D-) is the probability of not having the disease before the test, or 1 – P(D+)
P(T-|D-) is the specificity;
P(T-) is the probability of any negative test, that is, true negatives OR false negatives. Understanding this component: for a test to be a true negative, the disease must be absent, and the test must be negative in the absence of the disease: P(D-) x P(T-|D-). On the other hand, for a false negative test, the patient must have the disease P(D+), and the test must be negative in the patient with the disease P(T-|D+). Hence, the probability of a negative test is the sum of the probabilities of true and false negatives, as demonstrated below:

$$P(T-) = [(1-\ P(D+))\ x\ P(T-|D-)] + [P(D+)\ x\ P(T-|D+)]$$

where:

P(T-|D+) is the probability of a negative test given the disease present, which is the complement of sensitivity, that is: P(T-|D+) = 1 – P(T+|D+) or 1 – sensitivity

Rewriting the theorem, substituting the notations of conditional probability with the terms used in the diagnostic process, we obtain Bayes rule for the PPV. So:

$$P(D-|T-) = \frac{\big(1-P(D+)\big)\ x\ P(T-|D-)}{[(1-\ P(D+))\ x\ P(T-|D-)] + [P(D+)\ x\ (1-\ P(T+|D+))]}$$

becomes:

$$NPV = \frac{(1-prevalence)\ x\ specificity}{[(1-prevalence)\ x\ specificity] + [prevalence\ x\ (1-sensitivity)]}$$

6. Discussion

Beyond its role as a mathematical instrument, Bayes' rule is a rational framework for clinical decision, since it:

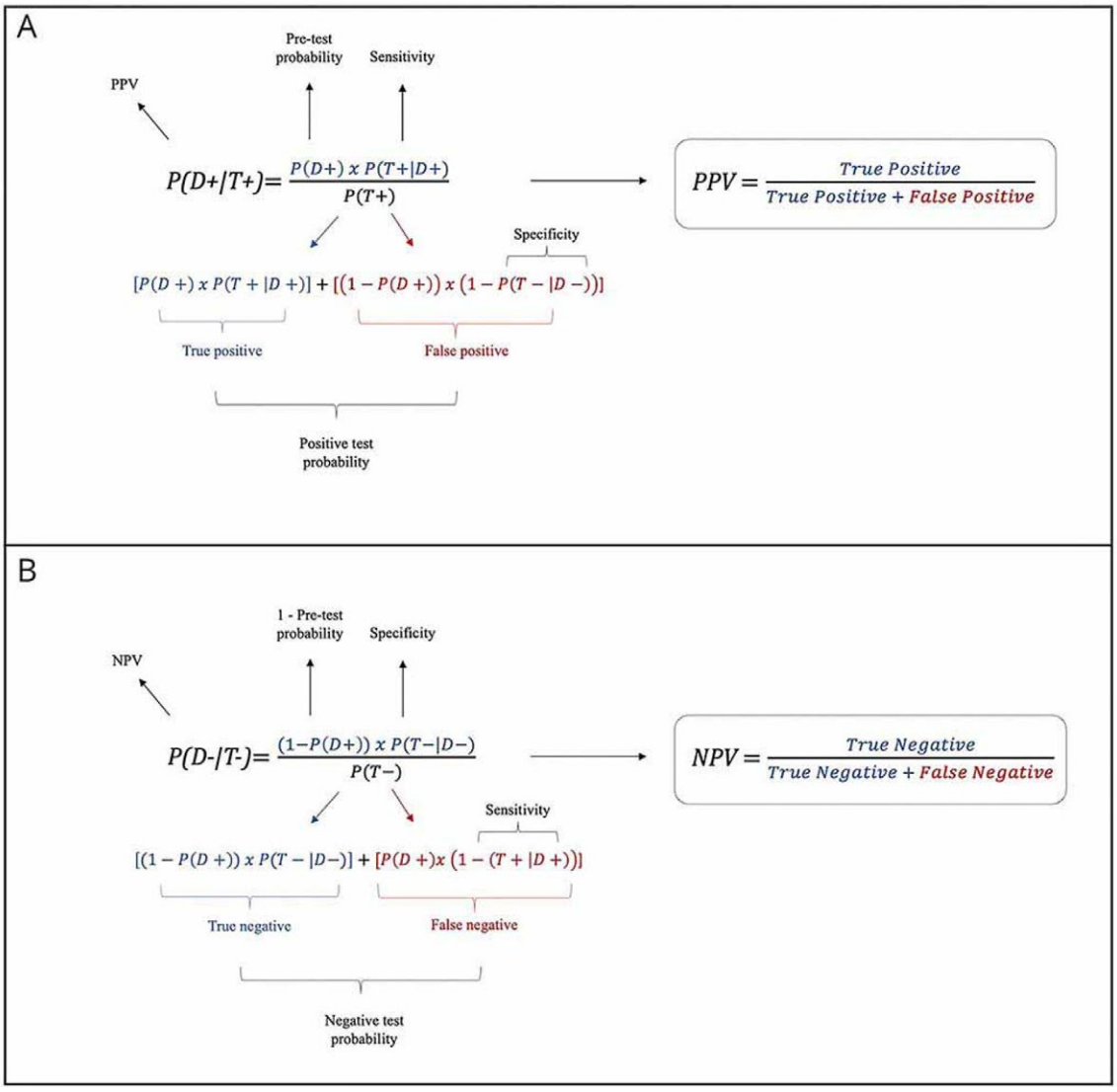
1. Isolates what truly matters for clinical decision-making (predictive values);
2. Mitigates the confusion between PPV and NPV with sensitivity and specificity, respectively, and demonstrates how these probabilities relate to each other;
3. Legitimizes clinical reasoning (represented by the pre-test probability) as the most important component of the diagnostic process.

We are not advocating the replacement of clinical judgment, which often involves a significant component of subjectivity[19], by cold numbers. Excellent doctors are masters at applying Bayesian reasoning without ever having heard of Thomas Bayes or using any formulas. They understand that the most important component of the diagnostic process is the reasoning that originates it and the resulting clinical decision depends on interpreting results considering their interaction with the context, not in isolation. Bad clinical reasoning results in bad clinical decisions, despite how accurate the diagnostic test: garbage in, garbage out[20]. Good clinical judgment extracts valuable information from the test, allowing clear communication of uncertainty with the patient and good shared decisions.

There are other ways to utilize the Bayesian framework such as applying likelihood ratios to pre-test odds[21,22], Fagan's nomogram[23,24], and natural frequency trees.[25] However, none of them works with the original structure of Bayes' theorem, which, in addition to the points listed above, works exclusively with probabilities, eliminating the need for conversion to odds and reconversion to probabilities, as with likelihood ratios.

We hope our step-by-step approach to Bayes' rule can help mitigate the dangerous confusion between sensitivity/PPV and specificity/NPV (Figure 2) and strengthen the idea that the value of a diagnostic test is directly proportional to the quality of clinical reasoning that leads to its request.

Figure 2. Bayes' theorem and its relationship with the components of the diagnostic process: Positive Predictive Value (A); Negative Predictive Value (B)



Source: the authors (2024).

Authors contributions

The authors declared that they have made substancial contributions to the work in terms of the conception or design of the research; the acquisition, analysis or interpretation of data for the work; and the writing or critical review for relevant intellectual content. All authors approved the final version to be published and agreed to take public responsability for all aspects of the study.

Conflicts of interest

No financial, legal, or political conflicts involving third parties (government, private companies, and foundations, etc.) were declared for any aspect of the submitted work (including but not limited to grants and funding, advisory board participation, study design, manuscript preparation, statistical analysis, etc.).

Indexers

The Journal of Evidence-Based Healthcare is indexed by DOAJ and EBSCO.

References

1. Morgan DJ, Pineles L, Owczarzak J, Magder L, Scherer L, Brown JP, et al. Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing. JAMA Intern Med. 2021;181(6):747-55. https://doi.org/10.1001/jamainternmed.2021.0269

2. White T, Algeri S. Estimating the lifetime risk of a false positive screening test result. PLoS One. 2023;18(2):e0281153. https://doi.org/10.1371/journal.pone.0281153

3. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2018;319(18):1914-31. https://doi.org/10.1001/jama.2018.3712

4. Lee JM, Arao RF, Sprague BL, Kerlikowske K, Lehman CD, Smith RA, et al. Performance of Screening Ultrasonography as an Adjunct to Screening Mammography in Women Across the Spectrum of Breast Cancer Risk. JAMA Intern Med. 2019;179(5):658-67. https://doi.org/10.1001/jamainternmed.2018.8372

5. Frank JW. Occult-blood screening for colorectal carcinoma: the risks. Am J Prev Med. 1985;1(4):25-32. https://doi.org/10.1016/S0749-3797(18)31396-5

6. Ismail AA. When laboratory tests can mislead even when they appear plausible. Clin Med (Lond). 2017;17(4):329-32. https://doi.org/10.7861/clinmedicine.17-4-329

7. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med. 2013;11(2):106-15. https://doi.org/10.1370/afm.1466

8. Morgan DJ, Brownlee S, Leppin AL, Kressin N, Dhruva SS, Levin L, et al. Setting a research agenda for medical overuse. BMJ. 2015;351:h4534. https://doi.org/10.1136/bmj.h4534

9. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ. 2002;324(7339):729-32. https://doi.org/10.1136/bmj.324.7339.729

10. Kassirer JP. Diagnostic reasoning. Ann Intern Med. 1989;110(11):893-900. https://doi.org/10.7326/0003-4819-110-11-893

11. Bayes T, Price R, Molina EC, Deming WE. An Essay towards solving a Problem in the Doctrine of Chances: Hafner; 1963. https://doi.org/10.1098/rstl.1763.0053

12. Bours MJ. Bayes' rule in diagnosis. J Clin Epidemiol. 2021;131:158-60. https://doi.org/10.1016/j.jclinepi.2020.12.021

13. Cosby K, Yang D, Fineberg HV. Assessing Diagnostic Performance. NEJM Evid. 2024;3(2):EVIDra2300232. https://doi.org/10.1056/evidra2300232

14. Manrai AK, Bhatia G, Strymish J, Kohane IS, Jain SH. Medicine's uncomfortable relationship with math: calculating positive predictive value. JAMA Intern Med. 2014;174(6):991-3. https://doi.org/10.1001/jamainternmed.2014.1059

15. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatr. 2007;96(3):338-41. https://doi.org/10.1111/j.1651-2227.2006.00180.x

16. Hoi A, Igel T, Mok CC, Arnaud L. Systemic lupus erythematosus. Lancet. 2024;403(10441):2326-38. https://doi.org/10.1016/s0140-6736(24)00398-2

17. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of Antinuclear Antibodies for Classifying Systemic Lupus Erythematosus: A Systematic Literature Review and Meta-Regression of Diagnostic Data. Arthritis Care Res (Hoboken). 2018;70(3):428-38. https://doi.org/10.1002/acr.23292

18. Li QZ, Karp DR, Quan J, Branch VK, Zhou J, Lian Y, et al. Risk factors for ANA positivity in healthy persons. Arthritis Res Ther. 2011;13(2):R38. https://doi.org/10.1186/ar3271

19. Norman GR, Brooks LR. The Non-Analytical Basis of Clinical Reasoning. Adv Health Sci Educ Theory Pract. 1997;2(2):173-84. https://doi.org/10.1023/a:1009784330364

20. Kilkenny MF, Robinson KM. Data quality: "Garbage in - garbage out". Health Inf Manag. 2018;47(3):103-5. https://doi.org/10.1177/1833358318774357

21. 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. https://doi.org/10.1111/j.1651-2227.2006.00179.x

22. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002;17(8):646-9. https://doi.org/10.1046/j.1525-1497.2002.10750.x

23. Fagan TJ. Letter: Nomogram for Bayes's theorem. N Engl J Med. 1975;293(5):257. https://doi.org/10.1056/nejm197507312930513

24. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. BMJ. 1997;315(7107):540-3. https://doi.org/10.1136/bmj.315.7107.540

25. Binder K, Krauss S, Schmidmaier R, Braun LT. Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. Adv Health Sci Educ Theory Pract. 2021;26(3):847-63. https://doi.org/10.1007/s10459-020-10025-8