

Does my patient have SARS-CoV-2 infection? A reminder of clinical probability formulas

Steven D Stovitz ,¹ Ian Shrier ²

10.1136/bmjebm-2020-111633

¹Department of Family Medicine and Community Health, University of Minnesota System, Minneapolis, Minnesota, USA

²Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, McGill University, Montreal, Québec, Canada

Correspondence to: **Dr Steven D Stovitz**, Family Medicine and Community Health, University of Minnesota System, Minneapolis, MN 55455, USA; stovitz@umn.edu

Introduction

Around the world, people are wondering, 'Do I have SARS-CoV-2?'. Determining a person's probability of illness is central to the practice and teaching of evidence-based medicine (EBM). Probabilities can be calculated through mathematical formulas, but studies find that many clinicians have difficulty with the calculations.¹⁻⁴ For many conditions in medicine, clinicians can rely on experience and/or established algorithms. However, with SARS-CoV-2 infection, clinicians lack experience and, given the rapidly changing prevalence of the illness, algorithms are difficult to establish. In this manuscript, we use the example of SARS-CoV-2 as an educational reminder of how clinicians can roughly gauge the probability of illness in different contexts and to highlight some conceptual issues in the estimates needed for the formulas.

Clinical diagnostic process and the formulas

Broadly speaking, the diagnostic process involves a sequence of estimating the probability of an illness and then adjusting this estimate with each new piece of information. Clinicians are generally taught to calculate the probability of illness given new information through either the use of Bayes probability theorem or the use of a 2×2 contingency table.

Mathematically, Bayes theorem and the 2×2 table are the same calculation. The 2×2 table calculation appears less complex than Bayes theorem, because it converts conditional probabilities into numbers that are entered into the cells of the table.²

Bayes probability theorem

In writing formulas, the term 'given' is symbolised as a straight vertical line. Letting 'P'=probability, 'A'=SARS-CoV-2 and 'B'=new information, Bayes probability theorem is: $P(A | B) = [P(A) \times P(B | A)] / P(B)$.⁵

Imagine that B, the new information, is the symptom of a cough. Then the probability of SARS-CoV-2 for a patient given the presence of cough, $P(A | B)$, is equal to the probability of SARS-CoV-2 in the patient's community, $P(A)$, multiplied by the probability of cough given that the patient has SARS-CoV-2, $P(B | A)$, all divided by the probability of cough in the patient regardless of their SARS-CoV-2 status, $P(B)$.

The 2×2 table calculation shown in [table 1](#) may be more familiar to clinicians as it is

often applied to laboratory tests where there is evidence of the accuracy of the test in those with and without the illness. The probability of illness given a positive test is termed the 'positive predictive value'. The probability of illness given a negative test is the complement of the 'negative predictive value'.

Clinical scenario

Consider a patient calling into your clinic with a report of a cough and asking, 'do I have SARS-CoV-2?'. To calculate the probability that this patient has SARS-CoV-2, you would need to estimate: (1) the prevalence of SARS-CoV-2 in the patient's community, (2) the probability of cough in those with SARS-CoV-2 and also either (3a) the probability of cough in the person regardless of SARS-CoV-2 infection (for Bayes theorem) or (3b) the probability of cough in the person if it were known that they did not have SARS-CoV-2 (for the 2×2 table). This calculated probability then represents the estimated prevalence for further calculations if new information about the patient is gathered, for example, a PCR test.

Assume that the patient is living in a community with a 10% prevalence of SARS-CoV-2, that the prevalence of cough in those with SARS-CoV-2 is 80%⁶ and that the probability of cough in the patient irrespective of SARS-CoV-2 is 20%. Furthermore, assume that the PCR test returns positive in 75% with SARS-CoV-2 (75% sensitivity) and returns negative in 95% without SARS-CoV-2 (95% specificity).⁷⁻⁹ To apply the formulas, any new information can be considered a test result, for example, the report of a cough can represent a 'positive cough test'.

The formulas with corresponding figures

In the figures, people are represented as yellow dots, those with SARS-CoV-2 are within the black background and those with a positive test are within the plaid background. The black area that overlaps with the plaid contains those who have SARS-CoV-2 and also test positive. [Figure 1](#) displays a 10% prevalence of SARS-CoV-2; among the 100 yellow dots, 10 (10%) are within the black background. Since 20% of the people have a cough, 20 of the 100 yellow dots are within the plaid background, and since the probability of cough in those with SARS-CoV-2 is 80%, 8 of the 10 in the black background are also within the plaid background. The formulas calculate a 40% probability of SARS-CoV-2 given a cough. This



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Stovitz SD, Shrier I. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2020-111633

Table 1 2x2 table calculating the probability of illness given test result

	Illness +	Illness -	Probability of illness given test result
Test +	True positive (TP)	False positive (FP)	$TP / (TP + FP) = PPV$
Test -	False negative (FN)	True negative (TN)	$FN / (FN + TN) = 1 - NPV$
	Number with illness	Number without illness	

NPV, negative predictive value; PPV, positive predictive value.

estimate can then be used as the prevalence if new information arises, for example, the result of a PCR test.

Figure 2 shows the probability of SARS-CoV-2 if this same person with a cough then receives a negative PCR test. To emphasise the sequential process of diagnostic reasoning, figure 2 has only 20 yellow dots representing the 20 people in our hypothetical population who had a cough. Since we estimated that the person with a cough had a 40% probability of SARS-CoV-2, 8 of the 20 yellow dots (40%) in figure 2 are within the black

background. Similar to many laboratory tests, the accuracy of the PCR test in those with and without the illness has been evaluated in studies. We are using evidence suggesting an approximate 75% sensitivity and 95% specificity.⁷⁻⁹ Therefore, among the eight with SARS-CoV-2, 75% (6) would have a positive PCR test and be placed inside the plaid background. Among the 12 without SARS-CoV-2, 95% (11) do not have a cough and lie outside of the plaid background. Both formulas show that the probability of SARS-CoV-2 in this person would now be about 15%. The BMJ

BAYES THEOREM AND THE 2X2 TABLE: PATIENT WITH COUGH: POSITIVE TEST

Assume:

- Prevalence of SARS-CoV-2 = 10%
- Probability of cough in those with SARS-CoV-2 (sensitivity) = 80%
- Probability of cough in the patient = 20% (calculated specificity = 87%)

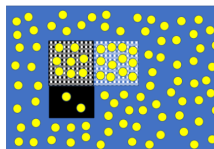
NEW INFORMATION = COUGH

Patient reports a cough (i.e. a "positive cough test")

VISUAL DISPLAY: POPULATION OF 100

Yellow dots = people

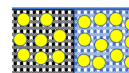
- Black background = Those with SARS-CoV-2
- Plaid background = Those with cough
- Overlap of Black and Plaid = Those with SARS-CoV-2 and cough



Probability of SARS-CoV-2 = 10%
(10% of 100 = 10)



Probability of cough = 20%
(20% of 100 = 20)



Among those with SARS-CoV-2
Probability of Cough, sensitivity = 80%
(80% of 10 = 8)

WHAT WE WANT TO KNOW

Among those who have a cough ,
what is the probability of SARS-CoV-2?

$$\frac{\text{8}}{\text{20}} = 40\%$$

CALCULATIONS

BAYES THEOREM CALCULATION

Let A = SARS-CoV-2
Let B = Cough
 $P(A | B) = [P(A) \times P(B | A)] / P(B)$
 $= [(10/100) \times (8/10)] / (20/100) = 8/20 = 40\%$

2X2 TABLE CALCULATION

	SARS-CoV-2	No SARS-CoV-2	Probability of SARS-CoV-2 given Cough +
Cough +	8	12	$8/20 = 40\%$
Cough -	2	78	
	10	90	Population = 100

Figure 1 The sensitivity of 80% was selected from a range of values in different studies.⁶ The prevalence of SARS-CoV-2 in the population and the probability of cough in the patient irrespective of SARS-CoV-2 status were chosen for educational purposes.

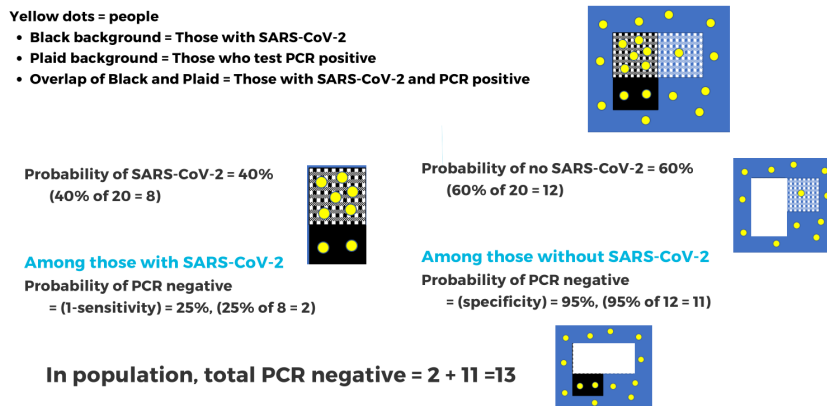
BAYES THEOREM AND THE 2X2 TABLE: SAME PATIENT WITH NEGATIVE PCR TEST

Assume:

- Prevalence of SARS-CoV-2 = 8/20 = 40% (calculated from Figure 1)
- Probability of PCR positive in those with SARS-CoV-2 (sensitivity) = 75%
- Probability of PCR negative in those without SARS-CoV-2 (specificity) = 95%

NEW INFORMATION = NEGATIVE PCR TEST

VISUAL DISPLAY: USING A POPULATION OF 20



WHAT WE WANT TO KNOW



CALCULATIONS

BAYES THEOREM CALCULATION

Let A = SARS-CoV-2
Let B = PCR negative
 $P(A | B) = [P(A) \times P(B | A)] / P(B)$
 $= [(8/20) \times (2/8)] / (13/20) = 2/13 = 15\%$

2X2 TABLE CALCULATION

	SARS-CoV-2	No SARS-CoV-2	Probability of SARS-CoV-2 given PCR -
PCR +	6	1	
PCR -	2	11	2/13 = 15%
	8	12	Population = 20

Figure 2 The prevalence of 40% was calculated in Figure 1. The sensitivity and specificity were selected from a range of values in different studies.⁷⁻⁹

has recently published an interactive online graphic that allows readers to input different estimates of pretest probability, sensitivity and specificity and then calculate the probability of illness given a test result.⁷

Some conceptual issues in the estimates needed for the formulas

The prevalence of the illness (here, SARS-CoV-2) in the patient's community

The prevalence or the 'pretest probability', symbolised as P(A) in Bayes theorem, is the probability that a person like our patient has the illness *prior to knowing* the additional information (test result). For many conditions in medicine, this estimate may be highly subjective. The current situation with SARS-CoV-2 is unique as many communities are receiving daily reports of community prevalence. Of course, if the patient has been exposed to someone with SARS-CoV-2 or is living in a specific area with a surge of

SARS-CoV-2 cases, then the pretest probability estimate should be higher than the prevalence in the larger community.

The probability of the new information (here, the report of cough or the negative PCR test) in those with the illness (here, SARS-CoV-2)

The probability of the new information (test result) occurring in those with the illness is symbolised as P(B|A) in Bayes theorem. The evidence for this probability comes from people with known illness. In the case of SARS-CoV-2, many with the virus are asymptomatic and may not be included in studies. Consequently, studies reporting the probability of a symptom in those with SARS-CoV-2 likely overestimates the probability of the symptom in all with SARS-CoV-2.

There is another reason why evidence may overestimate the probability of a symptom given an illness. Imagine that the patient with the cough in our example also denied having a fever. We

are now interested in the proportion with SARS-CoV-2 who have cough but are without fever, yet evidence from studies is rarely so granular.⁶ For example, the *US Naval Theodore Roosevelt* ship provides evidence of symptoms in non-hospitalised sailors with SARS-CoV-2.¹⁰ Approximately 35% of those with SARS-CoV-2 had a cough, and a similar proportion had a fever. Clearly, sailors with a cough and no fever represent a subset of the 35% sailors with SARS-CoV-2 who had a cough. So, relying solely on this information, the probability of a cough and no fever would be lower than 35%.

The probability of a positive test given illness is termed the sensitivity.¹¹ For the conditional probabilities of sensitivity and specificity (the probability of a negative test given no illness), clinicians are often taught the message of a 1975 publication that stated, 'For tests with binary outcomes, these measures are fixed'.¹² However, the measures are fixed *only* within a population similar to the particular patient and under the same testing circumstances.¹³ As an example, the sensitivity of the PCR test for SARS-CoV-2 differs depending on a variety of issues, such as time since exposure, symptoms and age.^{8 14}

The probability of the new information (here, the report of cough or the negative PCR test) in those without the illness (here, without SARS-CoV-2)

The probability of the new information in those without the illness is termed the specificity with a negative test result or the false-positive rate with a positive test result. This probability contributes to the denominator of both formulas, that is, P(B) in Bayes theorem. If the patient with a cough also had an underlying reason for coughing such as asthma, then the probability of cough without SARS-CoV-2 would be higher than if the patient had no underlying reason for developing a cough. With consistency in prevalence and sensitivity, the higher the probability of the new information in those without the illness, the lower the calculated probability of SARS-CoV-2 infection.

Conclusion

The current pandemic has clinicians around the world trying to gauge the likelihood that their patients have SARS-CoV-2. The purpose of this manuscript is to remind clinicians of two forms of a calculation to determine the probability of illness when presented with new information such as a patient symptom or the result of a laboratory test. The original Bayes probability theorem⁵ uses only proportional factors. The 2x2 contingency table calculation may be more familiar to clinicians and can be solved by converting the proportional factors of the Bayes formula into numbers for the cells of the table. The calculations require estimates that may be speculative and clinicians should recall that EBM recognises uncertainty, asking clinicians to use the 'best available evidence'.¹⁵ Understanding the calculations as well as the uncertainties inherent in some of the estimates needed for the calculations may help clinicians answer patient questions about SARS-CoV-2 as well as for more routine medical conditions.

Contributors SDS had the original idea for the manuscript. SDS and IS cowrote the manuscript and are responsible for the content.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iDs

Steven D Stovitz <http://orcid.org/0000-0002-4168-9823>

Ian Shrier <http://orcid.org/0000-0001-9914-3498>

References

- Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med* 1978;299:999–1001.
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al*. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8:53–96.
- Whiting PF, Davenport C, Jameson C, *et al*. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open* 2015;5:e008155.
- Stovitz SD. The inability to calculate predictive values: an old problem that has not gone away. *Med Sci Educ* 2020;30:685–8.
- Bayes T. An essay towards solving a problem in the doctrine of chances. 1763. *MD Comput* 1991;8:157–71.
- Struyf T, Deeks JJ, Dinnes J, *et al*. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database Syst Rev* 2020;7:CD013665.
- Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. *BMJ* 2020;369:m1808.
- Kucirka LM, Lauer SA, Laeyendecker O, *et al*. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* 2020;173:262–7.
- Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, *et al*. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS One* 2020;15:e0242958.
- Payne DC, Smith-Jeffcoat SE, Nowak G, *et al*. SARS-CoV-2 Infections and Serologic Responses from a Sample of U.S. Navy Service Members - USS Theodore Roosevelt, April 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:714–21.
- Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res* 1999;8:113–34.
- McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–5.
- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–30.
- Xiao AT, Tong YX, Gao C, *et al*. Dynamic profile of RT-PCR findings from 301 COVID-19 patients in Wuhan, China: a descriptive study. *J Clin Virol* 2020;127:104346.
- Sackett DL, Rosenberg WM, Gray JA, *et al*. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.