Journal Pre-proof

Why every clinician should know Bayes' rule

Bea Tiemens, Renée Wagenvoorde, Cilia Witteman

PII: S2452-3011(20)30046-8

DOI: https://doi.org/10.1016/j.hpe.2020.05.002

Reference: HPE 172

To appear in: *Health Professions Education*

Received Date: 30 April 2020

Revised Date: 3 May 2020

Accepted Date: 14 May 2020

Please cite this article as: Tiemens B, Wagenvoorde R, Witteman C, Why every clinician should know Bayes' rule, *Health Professions Education*, https://doi.org/10.1016/j.hpe.2020.05.002.

Why every clinician should know Bayes' rule[1]

Bea Tiemens[a, b, *], Renée Wagenvoorde[c] & Cilia Witteman[a*]

a. Behavioural Science Institute, Radboud University Nijmegen, The Netherlands

b. Pro Persona Research, Renkum, The Netherlands; Indigo Service Organization, Utrecht, The Netherlands

c. Faculty of Philosophy, Theology and Religious Studies, Radboud University Nijmegen, The Netherlands

* Corresponding author

Correspondence concerning this article should be addressed to Cilia Witteman, Behavioural Science Institute, Radboud University, P.O.Box 9104, 6500 HE Nijmegen, The Netherlands, C.Witteman@socsci.ru.nl

[1] This article is a translated and adapted version of: Tiemens B, Wagenvoorde R, Witteman, CLM. Waarom iedere psycholoog de regel van Bayes moet kennen. *GZ-Psychologie.* 2018;10:16-23. https://doi.org/10.1007/s41480-018-0034-7.

Abstract

We argue that knowledge about the rationale for Bayes' rule and about its proper application is a crucial tool for every clinician. We explain why such probabilistic reasoning is so important. We then acknowledge that the rule is quite abstract and may be difficult to use, and we offer a guideline to overcome this difficulty. We illustrate our text with an example from the mental health domain, but indicate that the same reasoning applies to every clinical area. Finally, we give a five-step recipe for interpreting test results properly.

Keywords: probabilistic reasoning, Bayes, sensitivity and specificity, interpreting test results

1. Introduction

All clinicians, both medical doctors and clinical psychologists, have been taught Bayes' rule at university, but few remember it exactly. "Wasn't that one of those complicated formulae? Oh well, it has nothing to do with clinical practice anyway, so no matter." But beware; Bayes' rule is extremely relevant to clinical practice. You ignore it at your peril. We explain the rule and its relevance with an example and we show how easy it is to use it.

Example:

Suppose you work in a psychiatric institute and a patient is referred to you by their general practitioner (GP) because of an elevated score on a depression questionnaire. In this GP practice, 11% of the patients have a depression (the 'prevalence' of the disorder). If a patient actually has a depression, the likelihood that they have a positive score on this depression questionnaire is 83% (the 'sensitivity' of the test). A patient who does not have a depression, has an 80% chance of a negative score (the 'specificity' of the test). What is your estimate of the probability that the referred patient has a depression?

All clinicians have been taught how to answer this question. Indeed, you need to consider the sensitivity and the specificity of the depression test, but you should also take into account the prevalence of the condition tested for. But how? Many people will reason as follows: "I know what I should do: enter the numbers into Bayes' rule. But that was quite complicated, so let's see: if this patient has a high score on a depression questionnaire, obviously the likelihood of their having a depression is fairly high too, right? You don't have to do the maths, and certainly not using such a formula that after all has nothing to do with

clinical practice." Bayes' rule, however, is extremely relevant to clinical practice. When you ignore the rule, you can make serious errors. This may actually happen often. For example, much has been written about underdiagnosing depressive disorders by GPs. In a meta-analysis GPs were found to miss 50% of their patients' depressive disorders and they did diagnose a depression in almost 20% of the patients without a depression.[1] In absolute numbers far more patients were overdiagnosed (and probably treated) than underdiagnosed, because there are far more patients who are not depressed. Overdiagnosis by not following Bayes' rule properly may thus lead to overtreatment. In a classical study on physicians who needed to decide whether or not to perform a biopsy on women with possibly malignant breast masses, it was shown that physicians were not very good at the probabilistic reasoning required to make this decision, with too many interventions as a consequence.[2] Overdiagnosis of breast cancer has been found to be as high as 52%.[3] Overdiagnosis is also found to occur with prostate cancer, leading to overtreatment.[4] The same will be true of other medical conditions when the extra step of taking the base rate of that condition into account is not properly followed.

We will explain Bayes' rule and its importance with the help of the depression example given above, both by applying Bayes' rule itself and also in two simpler ways.

2. Bayes' rule in words

Bayes' rule is designed to help calculate the posterior probability of an individual's condition (disease or disorder) from three elements: the prevalence or prior probability of that condition in the population of that individual, the outcome of a test and the psychometric characteristics of that test: its sensitivity and specificity. To use Bayes' rule it is essential to know the frequency of occurrence of the condition in the population of which the individual is a member. This is the element that is often omitted, leading to erroneous conclusions. If in a given population a condition is rare (that is: the prevalence of that condition is low), the

probability that the individual has the condition can never be high, even though it is higher with a positive test result than before the test outcome was known. With a high prevalence, the probability of the condition is high before testing, and higher with a positive test result. This means that in different populations, where the prevalences of the condition is different, the same test result leads to a different probability of the same condition. This is intuitively difficult to follow, even if you are familiar with Bayes' rule and with probabilistic reasoning. Maybe it becomes easier when you consider, for example, testing for a tropical disease either in the tropics or in The Netherlands. In The Netherlands the probability that a positive test result is a pure coincidence is much higher, because the prevalence of the tropical disease is much lower.

Below we show, using our depression example, that Bayes' rule might seem difficult, but that the calculation is actually childishly simple.

2.1 Does the patient have a depression?

In our example, the prior probability that a random patient seen by the GP has a depression is 11%. This is the prevalence (prev) of depression in the GP's population, also called the 'base rate' or the pre-test probability. The left arm of the left panel of figure 1 shows this 11% probability of depression. Below are shown the 83% sensitivity of the test: the probability that a patient who actually has a depression, has a positive score on a depression questionnaire. In the right arm the 80% specificity of the test is shown: the probability that a patient who does not have a depression, scores negatively. These probabilities are based on research by Cameron and colleagues (2011) who tested the Brief Depression Inventory II (BDI-II) in GP practices.[5]

Insert Figure 1. about here

Bayes' rule is shown in the right panel of Figure 1. The P stands for 'probability' or likelihood. The vertical line | means 'under the condition that' or 'given'. So P(D+|T+) means the probability (P) that the patient has a depression (D+) given (|) a positive test result (T+) on a depression questionnaire. This is the probability we want to know, and we can use Bayes' rule to calculate it, using the following elements:

P(T+|D+) is the probability that a patient with a positive test result has a depression, so the sensitivity of the test, here 0.83.

P(D+) is the probability that a patient in this population has a depression, the prevalence, here 0.11.

P(T+|D-) is the probability that a patient with a positive test has no depression, i.e. 1-specificity (the probability that a patient with a negative score on the test does not have depression), here 1-0.8 = 0.2

P(D-) is the probability that a patient in this population has no depression, 1-prevalence, here 1-0.11 = 0.89.

So, we have all the information required to complete Bayes' rule:

P(D+|T+) = (P(T+|D+) x P(D+)) / (P(T+|D+) x P(D+)) + (P(T+|D-) x P(D-)) =

(0,83 x 0,11) / (0,83 x 0,11) + (0,20 x 0,89) = 0,34

Thus, the probability (P) that the patient from this general practice population has a depression (D+) given (|) a positive test result (T+) on a depression questionnaire, P(D+|T+), is about one third, 34%. This is also called the posterior or post-test probability; the probability after the test has been taken and given the outcome of the test.

3. Bayes made easier

Applying Bayes' rule requires some puzzling, and the formula is hard to remember. It becomes much easier if the same information is presented in absolute numbers instead of in probabilities,[6] as is done in Figure 2.

Insert Figure 2. about here

On the left the information is in the same flowchart as in figure 1 and on the right in a cross tabulation, but now in absolute numbers. Using the sensitivity of the test we are able to calculate how many patients with a depression will have a positive score: 0.83 x 110 = 91.3, rounded to 91. So 110 - 91 = 19 patients with depression will have a negative score on the depression test, and thus their disorder will be missed. The specificity is used to calculate how many patients who do not have a depression indeed score negatively on the test: 0.80 x 890 = 712. So, 890 - 712 = 178 patients who do not have a depression will have a positive score on the depression test, and will be overdiagnosed. Presented like this it immediately becomes clear how many patients have a positive test score (91 + 178) and how many of those actually have a depression (91). The probability that a patient has a depression given a positive test result is therefore: 91/91+178 = 0,34. The probability that the patient does not have a depression given a negative test result is: 712/19+712 = 0,97

4. Do not rely on sensitivity and specificity alone

Although the sensitivity of the BDI in this study[3] is 83% and the specificity 80%, the probability that we are interested in, the posterior probability that the patient with a positive test result actually has a depression, is much lower, namely only a third (34%). This is much too low for a treatment indication. A negative test result, on the other hand, gives a relatively

high level of certainty about the absence of depression, indeed a probability that is higher than the specificity. This test could therefore justifiably be used to exclude depression, but not to establish its presence.

From the above it is obvious that you should not simply rely on the score of a test. A positive test result does not necessarily mean that the condition is present and a negative result does not necessarily indicate that it is absent. Neither can we rely on the sensitivity or 'chance of being caught' and the specificity of an instrument only. Especially when the prevalence of the condition is very low, a positive test result will never translate to a very high probability of the condition even when the sensitivity and specificity of the test are good.

This becomes clear when we use the same depression test, with the same sensitivity and specificity, but now in the general population, where the prevalence of depression is lower than in general practice: only 5%. Then the probability of depression with a positive test result is less than 20% (figure 3): 42 / (42 + 190) = 0.18. Suppose that in the situation of figure 3 the specificity is not 80% but 95%, then the number of non-depressed patients with a positive score will be 48. The post-test probability is much higher now, 42 / (42 + 48) = 0.47, but is still lower than 50%.

Insert Figure 3. about here

5. Research versus practice

The examples show that it is necessary to always keep Bayes' rule in mind in clinical practice and apply it when interpreting test results. When the sensitivity and specificity of a test are established in scientific research, this is based on knowledge of how many respondents do and how many do not have the condition. In clinical practice this is unknown, and that is precisely why the test is administered. A myopic clinician then only sees a positive test result. In the

above example, 23% (42 + 190/1000) of the people tested have a positive test result, while only 5% actually have a depression. The clinician does not know which of those 23% do or do not have a depression. But if the clinician is aware of Bayes' rule, it is clear that extra diagnostics are needed in addition to administering a BDI.

Bayes' rule, in the formula or in one of the other simpler forms, helps determine the posterior probability of a condition, given three elements: the test result, the pre-test or prior probability (the prevalence) and the sensitivity and specificity of the diagnostic test. Clinicians often forget to include the pre-test probability in their interpretation; if they do look further than the test result, they tend to only consider the sensitivity of a test. Typically, a positive test score is interpreted as the presence of the condition. How easy to draw but how incorrect this conclusion is, should be clear by now.

6. Bayes in clinical practice

When clinicians are aware of Bayes' rule, this means that when interpreting a test result they will take into consideration the setting in which the test was taken and the population to which the tested person belongs, and they will establish the prevalence of the condition in that setting and population. The prevalence of the most common mental illnesses in the general population can be found in population studies,[7] e.g., for The Netherlands in the Nemesis study.[8] Statistics about different types of cancer can also be found online, e.g., for the UK in publications of the Cancer Research UK (https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type). Most clinicians however, do not work at the general population level, and the prevalence in their practice or institute is more relevant.

Most numbers on prevalence in general practice are based on registrations by the GP. To stay with our depression example: We know that GP registrations are generally an

underestimation.[1] The prevalence mentioned in the Dutch general practice guidelines (https://www.nhg.org/standaarden/volledig/nhg-standaard-depressie-tweede-herziening#idm10758112) is, for example, lower than the prevalence at population level. Because almost everyone in the Netherlands has a GP, the prevalence in general practice should be comparable to the prevalence at the population level. However, people with a depression visit their GPs more often than people without depression. Thus the prevalence of depression is higher among those who come to the GP's office for consultation. The same will be true of any other condition. A rough guideline is to double the population prevalence.

The prevalence of a mental illness in your work setting is best determined on the basis of registrations in your own organization. Suppose that about a third of the patients who are referred to your organization have a depression. If you enter that prevalence in the flowchart as in figure 2 and keep the same sensitivity and specificity, the probability of depression of a random patient with a positive score on the BDI is about two thirds.

7. Conclusion

Ignoring Bayes' rule leads to overdiagnosis of conditions that do not occur so often (with a low pre-test probability). This is easily understood by looking at the flowcharts. Because with low prevalence the 'no disorder or disorder' group is very large, the post-test probability will remain low even with very high specificity, despite the large numbers of people who have a depression and who tested positive. Making the error of omitting to take prevalence into account can have major consequences for the treatment policy and thus the course of a patient's symptoms. We therefore strongly recommend remembering at least the gist of Bayes' principle: given a positive test result, the probability of a condition depends on the population and setting of the patient, on the prior probability of their having the condition. As shown above, the calculation is easy to do with absolute numbers. The steps are summarized

in the box below. Calculating posterior probabilities using a group of 100 or 1000 patients is very instructive and provides quick insight into the probability of a condition given a score on a specific test.

Steps in interpreting a test result:

1. Determine the population to which the person tested belongs and the setting in which the test was taken
2. Find the prevalence of the condition in that population and setting
3. Find the sensitivity and specificity of the test (in the manual of the test)
4. Enter the numbers in a flowchart or crosstable as in figure 2
5. Calculate the probability of the presence of the condition given a positive test result, or the probability of the absence of the disorder given a negative test result.

References

1. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374(9690):609-19. doi:10.1016/S0140-6736(09)60879-5

2. Eddy DM. Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press; 1982:249-267. doi:10.1017/CBO9780511809477.019

3. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: Systematic review of incidence trends. *BMJ* 2009; 339:b2587 doi.org/10.1136/bmj.b2587

4. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol*. 2014;65(6):1046□1055. doi:10.1016/j.eururo.2013.12.062

5. Cameron IM, Cardy A, Crawford JR, et al. Measuring depression severity in general practice: Discriminatory performance of the PHQ-9, HADS-D, and BDI-II. *Br J Gen Pract.* 2011;61(588):e419-e426. doi:10.3399/bjgp11X583209

6. Gigerenzer G. What are natural frequencies? *BMJ.* 2011;343:d6386. doi:10.1136/bmj.d6386

7. Steel Z, Marnane C, Iranpour C, et al. The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. *Int J Epidemiol.* 2014;43:476-493. doi:10.1093/ije/dyu038

8. De Graaf R, Ten Have M, Van Gool C, et al. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol.* 2012;47:203-213. doi:10.1007/s00127-010-0334-8

Figure Captions

Figure 1: The probability of depression with a positive score on a depression test: Bayes' formula

Figure 2: The probability of depression with a positive score on a depression test, in numbers

Figure 3: The probability of depression with a positive test result in the general population

Figure 1.

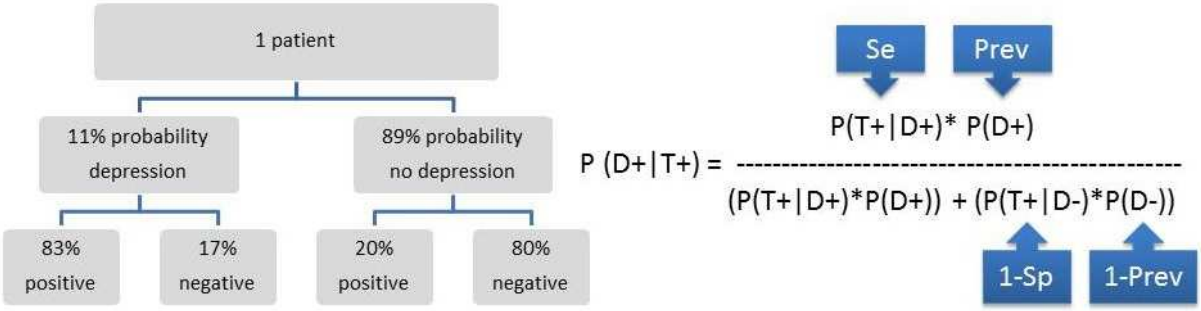
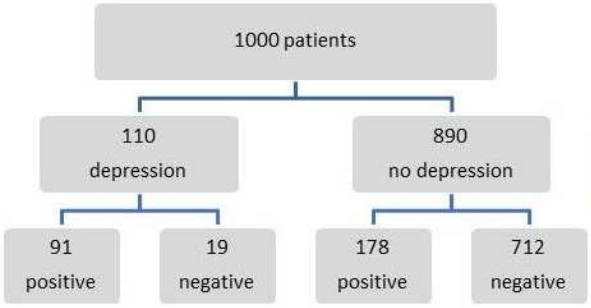


Figure 2



		Depression		
		present	absent	
Test	positive	91	178	269
	negative	19	712	731
		110	890	1000

Figure 3