

# The Benefits of Applying Bayes' Theorem in Medicine

David Trafimow<sup>1</sup>

Department of Psychology, MSC 3452  
New Mexico State University, P. O. Box 30001  
Las Cruces, NM 88003-8001

**Abstract:** The present article provides a very basic introduction to Bayes' theorem and its potential implications for medical research. This introduction was written to be accessible to medical researchers without much mathematical background in general or without much background in Bayesian mathematics specifically. I prove Bayesian equations from very basic probability theorems and also show how these basic theorems can aid medical researchers to make important discoveries without having to do much empirical work. In many cases, the necessary empirical work, including relevant epidemiological work, is already in the medical literature and important discoveries can be made merely by looking up that work and applying one of the Bayesian equations to be presented. I also demonstrate that the traditional null hypothesis significance testing procedure that currently dominates medical research is blatantly invalid; a Bayesian point of view suggests that researchers should not use this procedure. Finally, I discuss some special issues, including philosophical issues that pertain to using a Bayesian perspective in medicine.

## I. INTRODUCTION TO THE BAYESIAN SPECIAL ISSUE

Any researcher who is knowledgeable about the famous theorem by Bayes cannot help but notice two unfortunate patterns when perusing the medical findings. First, there already is much valuable published information but researchers are unable to use that information to maximal advantage. For example, there have been countless epidemiological studies that provide the medical world with potentially useful base rate information about the prevalence of diseases. By combining epidemiological information with false positive and false negative rates gleaned from research on particular medical diagnostic procedures, Bayesian analysis renders it possible to compute important probabilities, such as the probability that a person has a disorder given the result of the diagnostic procedure. Critical discoveries such as this can be made simply by spending an hour finding the relevant studies, and spending three minutes using Bayes' theorem to make the relevant computation. An important goal of the present article, and of this Special Issue more generally is to make accessible to medical researchers (a) that Bayesian analyses greatly expand the potential for making important discoveries in medicine and (b) that some Bayesian analyses are very easy to perform and even can be performed on the backside of an envelope.

The second problem with medical researchers being unfamiliar with Bayesian analyses is that they depend too much on the null hypothesis significance testing procedure ( $p < .05$ ). The procedure is flawed in basic ways that are obvious to a Bayesian but that are difficult to see from the point of view of a non-Bayesian. The organization of the present article will underscore both issues. I commence with a general introduction of Bayes' theorem that should be easy for any medical researcher who has had rudimentary algebra to understand. Based on the introduction, the following sections discuss, in turn, how to exploit Bayes' theorem even if one has very little mathematical background and the implications of the theorem for null hypothesis significance tests. Finally, there are auxiliary issues that remain to be discussed.

## II. INTRODUCTION TO BAYES' THEOREM IN A MEDICAL CONTEXT

Suppose that a researcher wishes to know the probability of a disorder given the result of a diagnostic procedure or test. I will use the following notation.

- $P(D)$ : Base rate probability (proportion of cases in the relevant population) of the disorder.
- $P(\sim D)$ : Proportion of cases in the relevant population that do not have the disorder [ $P(\sim D) = 1 - P(D)$ ].
- $P(T)$ : Probability of a positive result when a test or diagnostic procedure is performed.
- $P(\sim T)$ : Probability of a negative result when a test or diagnostic procedure is performed [ $P(\sim T) = 1 - P(T)$ ].

---

<sup>1</sup> Corresponding Author: [dtrafimo@nmsu.edu](mailto:dtrafimo@nmsu.edu)

- $P(D|T)$ : Conditional probability that the patient actually has the disorder given that a positive result on the test. This is usually what the physician or medical researcher wishes to know.
- $P(T|D)$ : Conditional probability that the test will come out positive given that the patient has the disorder. This also can be considered to be the true positive rate.
- $P(T|\sim D)$ : Conditional probability that the test will come out positive given that the patient does not have the disorder. This also can be considered to be the false positive rate.
- $P(\sim T|D)$ : Conditional probability that the test will come out negative given that the patient has the disorder. This also can be considered to be the false negative rate.
- $P(\sim T|\sim D)$ : Conditional probability that the test will come out negative given that the patient does not have the disorder. This also can be considered to be the true negative rate.

Using the foregoing notation, Bayes' theorem can be expressed as Equation 1 below and gives the conditional probability that the patient has the disorder given that a positive test result has been obtained.

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

However,  $P(T)$  is usually not known. So a generally more useful form of the theorem can be expressed as Equation 2 below.

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(T|\sim D)[1 - P(D)]} \quad (2)$$

Equations 1 and 2 might have seemed to come out of nowhere. In the following two sections, I prove Equations 1 and 2 from a basic probability theorem. Those who are willing to accept Equations 1 and 2 without proof can skip these sections.

### **Proof of Equation 1**

I commence with a basic probability theorem expressed below as Equation 3. The basic idea is that the probability of joint events equals the probability of one of the events multiplied by the conditional probability of the other event, given the first event.

$$P(D \text{ and } T) = P(D)P(T|D) \quad (3)$$

The same rule that generates Equation 3 also generates Equation 4.

$$P(T \text{ and } D) = P(T)P(D|T) \quad (4)$$

Clearly,  $P(D \text{ and } T) = P(T \text{ and } D)$  and so Equation 5 follows directly.

$$P(T)P(D|T) = P(D)P(T|D) \quad (5)$$

Dividing both sides of Equation 5 by  $P(T)$  provides one form of Bayes' theorem. This is Equation 1 written again for the reader's convenience.

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

### **Proof of Equation 2**

As I pointed out earlier, researchers often do not have direct access to  $P(T)$ , which renders Equation 1 difficult to implement. However, it is possible to partition  $P(T)$ , which is the denominator of Equation 1, into additional variables. Equation 6 below provides an obviously valid starting point for this partitioning.

$$P(T) = P(T \text{ and } D) + P(T \text{ and } \sim D) \quad (6)$$

The same rule that lead to Equation 3 and Equation 4 also leads to Equation 7, which converts  $P(T \text{ and } D)$  from Equation 6 into a product of probabilities.

$$P(T \text{ and } D) = P(D)P(T|D) \quad (7)$$

Using the rule again on  $P(T \text{ and } \sim D)$  from Equation 6 leads to Equation 8.

$$P(T \text{ and } \sim D) = P(\sim D)P(T|\sim D) \quad (8)$$

I add Equations 7 and 8 to get Equation 9.

$$P(T \text{ and } D) + P(T \text{ and } \sim D) = P(D)P(T|D) + P(\sim D)P(T|\sim D) \quad (9)$$

But going back to Equation 6, I replace  $P(T \text{ and } D) + P(T \text{ and } \sim D)$  with  $P(T)$  which yields Equation 10.

$$P(T) = P(D)P(T|D) + P(\sim D)P(T|\sim D) \quad (10)$$

Now, referring back to Equation 1, Equation 10 is used to replace the denominator of Equation 1. That is,  $P(T)$  is replaced with  $P(D)P(T|D) + P(\sim D)P(T|\sim D)$  to yield Equation 11.

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(\sim D)P(T|\sim D)} \quad (11)$$

Although Equation 11 is useful in its own right, it contains four probabilities on the right side of the equation. These are  $P(D)$ ,  $P(T|D)$ ,  $P(T|\sim D)$ , and  $P(\sim D)$ . It can be convenient to reduce these four probabilities to three probabilities by eliminating  $P(\sim D)$ . This can be accomplished easily by using Equation 12.

$$P(\sim D) = 1 - P(D) \quad (12)$$

Replacing  $P(\sim D)$  with  $1 - P(D)$  in Equation 11 yields Equation 2, written again for the reader's convenience.

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(T|\sim D)[1 - P(D)]} \quad (2)$$

### III. USING BAYES' THEOREM TO EXPLOIT EPIDEMIOLOGICAL DATA

The vast wealth of published epidemiological findings provides researchers with the ability to make important discoveries just by looking up relevant data and using Equation 2. For example, consider some recent work on using the results from MRI scans to diagnose Herniated Nucleus Pulposus (HNP). Although many physicians place strong dependence on MRI scans to diagnose HNP, it is not clear that it is an optimal diagnostic strategy. A hint that the strategy might not be optimal comes from researchers who have found substantial false positive rates [1, 2]. Ultimately, to assess the validity of using positive MRI results to diagnose HNP, it is necessary to know the probability that a patient has the disorder given that a positive MRI result has been obtained. Based on Equation 2, it is obvious that knowledge of three probabilities is essential: The base rate probability of HNP, the probability of a positive MRI result given that the patient has the disorder, and the probability of a positive MRI result given that the patient does not have the disorder (the false positive rate). Fortunately, all of this information is available in the literature. Remaining consistent with the nomenclature I have been using, these probabilities are as follows:  $P(D) = .030$  [3, 4, 5, 6],  $P(T|D) = .804$  [7], and  $P(T|\sim D) = .300$  [1, 2]. The probabilities can be plugged into Equation 2 as follows [8].

$$P(D|T) = \frac{(.030)(.804)}{(.030)(.804) + (.300)(1 - .030)} = .077$$

Stated verbally, the probability that a patient has HNP given a positive MRI result is less than 8%. In turn, this implies that the probability that the patient does not have HNP given a positive MRI result is greater than 92%.

The foregoing example highlights certain points about opportunities that a Bayesian perspective offers medical researchers that that I wish to emphasize. First, the valid use of Equation 2 depends, in large part, on having good base rate information; that is, the researcher needs to be able to obtain  $P(D)$ . As opposed to many other sciences, medicine has a particular advantage in this respect. Many other sciences do not have good base rate information available because they do not have the equivalent of epidemiological studies. Perforce, medicine is better suited than many other sciences to exploit the full power of Bayesian inference. As the example demonstrates, it often is possible to make important discoveries merely by looking up the relevant epidemiological information and using Bayes' Theorem.

Second, although the presence or absence of relevant epidemiological findings sets medicine apart from other sciences in the Bayesian context, it also is necessary to know false positive and false negative rates [ $P(T|D)$  and  $P(T|\sim D)$ , respectively]. Like base rate information, false positive and false negative rates often can be obtained from the literature. In those cases where false positive and false negative rates are not available, there is an opportunity for medical researchers to make important empirical contributions by performing research to obtain them.

#### IV. NULL HYPOTHESIS SIGNIFICANCE TESTING

Like many sciences, medical research depends to a large extent on the null hypothesis significance testing procedure (NHSTP). The NHSTP includes, at a minimum, the following steps [9].

- The researcher proposes a substantive hypothesis to be (hopefully) supported.
- The researcher proposes an alternative null hypothesis to be (hopefully) rejected in favor of the substantive hypothesis.
- The researcher collects the data.
- The researcher computes the probability of obtaining the finding (or one more extreme), given the null hypothesis. This often is symbolized as  $p$ .
- If  $p$  is less than .05, the researcher rejects the null hypothesis in favor of the substantive hypothesis. The shorthand for saying that the researcher was able to reject the null hypothesis in favor of the alternative hypothesis, based on  $p < .05$ , is that the finding was “statistically significant” or simply “significant.”

But there is a logical problem with Steps 1-5 [9]. To understand the problem, consider the goal, which is to reject the null hypothesis in favor of the substantive hypothesis. What probability does the researcher need to know to have a valid basis for rejecting the null hypothesis? The obvious answer is that the researcher needs to know the probability of the null hypothesis, given the finding. Unfortunately,  $p$  does not provide that. Rather,  $p$  is the probability of obtaining a finding (or one more extreme than the finding actually obtained), given that the null hypothesis is true. As the great statistician Fisher explained at length [10], the probability of getting the finding given that the null hypothesis is true, is the inverse conditional probability of the one that is needed which, again, is the probability of the null hypothesis given the finding. The logical problem is stated succinctly in Equation 13, where  $P(F|H_0)$  denotes the probability of the finding given the null hypothesis (or  $p$ ) and  $P(H_0|F)$  denotes the probability of the null hypothesis given the finding. Note the inequality sign in Equation 13 that demonstrates the fallacy of computing  $p$  and believing that it represents the probability of the null hypothesis given the finding.

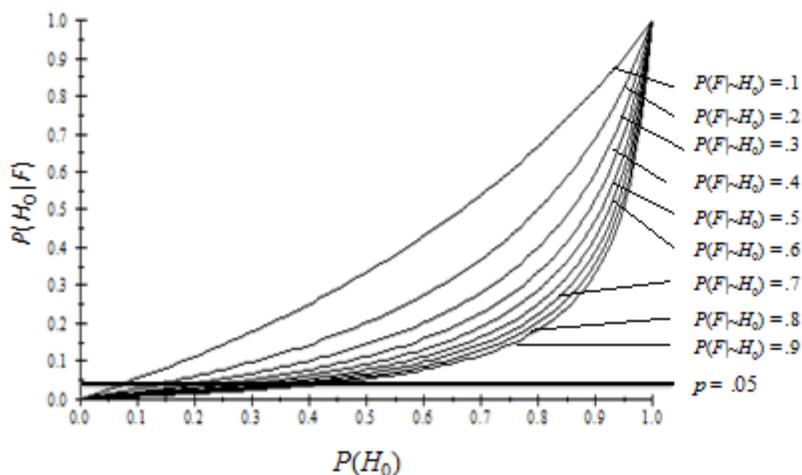
$$p = P(F|H_0) \neq P(H_0|F) \tag{13}$$

The relationship between the probability of the finding given the null hypothesis and the probability of the null hypothesis given the finding is provided by Bayes’ Theorem, as Equation 14 shows, where  $P(H_0)$  denotes the base rate probability of the null hypothesis (note the similarity to Equation 2).

$$P(H_0|F) = \frac{P(H_0)P(F|H_0)}{P(H_0)P(F|H_0) + P(F|\sim H_0)[1 - P(H_0)]} \tag{14}$$

Equation 14 makes clear that what we really need to know to reject the null hypothesis, which is  $P(H_0|F)$ , is a function of three variables of which  $P(F|H_0)$  (or  $p$ ) is only one. The other two variables are the base rate probability of the null hypothesis (often termed the “prior” probability of the null hypothesis), and the probability of the finding given that the null hypothesis is not true. Based on Equation 14, I have shown that under most values these latter two probabilities can take on, even keeping  $p$  constant at the conventional level of .05, the probability of the null hypothesis given the finding tends to be well above .05 [9]. I have recreated the figure as Figure 1, where a line going across the lower part of the figure represents the conventional .05 level.

Figure 1 demonstrates three Bayesian implications. First, much of the curves are above the line, thereby indicating that  $p$  often underestimates the probability of the null hypothesis given the finding. Second, to get the probability of the null hypothesis given the finding below the dashed line (so that there is some validity to the NHSTP), it is necessary to have a low prior probability of the null hypothesis. But if the prior probability of the null hypothesis is low, then very little is learned from the NHSTP because it results in the rejection of a hypothesis that is already known to be wrong; validity and information gained go in opposite directions. That is, as the validity of the NHSTP increases (because the prior probability of the null hypothesis decreases), less information is gained by rejecting that which is already known to be false. This is an obviously undesirable state of affairs. Finally, as the probability of the finding given that the null hypothesis is not true increases the probability of the null hypothesis given the finding decreases. Based on Equation 14, Figure 1 illustrates the negative mathematical consequences of using the NHSTP. There are much more extensive discussions that bring out additional problems [9, 11, 12, 13, 14].



**Fig1.** The posterior probability of the null hypothesis [ $P(H_0|F)$ ] as a function of its prior probability [ $P(H_0)$ ] and the probability of the finding given that the null hypothesis is not true [ $P(F|\sim H_0)$ ]. The prior probability of the null hypothesis ranges from 0 to 1 and the probability of the finding given that the null hypothesis is not true takes on values of .1, .2, .3, .4, .5, .6, .7, .8, and .9. The solid black line represents  $p$ , which has been set at the conventional alpha level of .05.

Although many statisticians have recognized the basic logical invalidity of the NHSTP [9, 11, 12, 13, 14] it continues to be used in most sciences, including medicine. Some have tried to excuse the invalidity by on the one hand admitting the logical problem while on the other hand claiming that it does not matter too much because  $p$ -values are strongly correlated with probabilities of null hypotheses [15]. Because of this strong correlation, the argument goes,  $p$  provides a strong basis for rejecting the null hypothesis even though the inference is logically invalid. In one sense, the argument has some basis. As I showed above in Equation 14,  $p$  is one of three variables that determine the probability of the null hypothesis given the finding and so  $p$  has to be correlated with  $P(H_0|F)$ . Nevertheless, there is no reason to believe that the size of the correlation is sufficiently impressive to justify the NHSTP. In fact, I showed that the correlation is small ( $r = .396$ ) so that the variance in probabilities of null hypotheses given findings accounted for by variance in  $p$ -values is less than 16% [16]. Worse yet, the practice in most sciences, including medicine, is to dichotomize  $p$ -values by having a cut-off at .05, often termed the alpha level. The use of an alpha level greatly reduces the correlation even from the small number that we already have seen. In the case where alpha equals .05, the correlation reduces to .289 and so the variance in probabilities of null hypotheses given findings accounted for by  $p$ -values is less than 9%, thereby leaving over 91% unaccounted for! Thus, the correlation between  $p$  and the probabilities of null hypotheses given findings is blatantly insufficient to justify using  $p$  to draw conclusions about the probability of the null hypothesis given the finding.

The usefulness of Bayes' theorem is not limited to its ability to demonstrate the invalidity of the NHSTP and the negative mathematical consequences of that invalidity. It also can be used to test competing medical hypotheses. To see that this is so, imagine that Physician 1 hypothesizes that a particular medical procedure will result in a cure 50% of the time whereas Physician 2 believes that the procedure will result in a cure 65% of the time. They test 20 patients who have the relevant disorder and obtain 13 cures out of 20 attempts. If they used the NHSTP to solve the disagreement, they would perform a significance test to find the probability of 13 successes out of 20 tries based on the null hypothesis of a 50% cure rate (as in a coin toss), and the result would be that  $p = .26$  (using the usual two-tailed test). This result is not even close to the conventional alpha level of .05 and so the null hypothesis would be retained, thereby supporting the hypothesis of Physician 1 over the hypothesis of Physician 2. But we have already seen that the NHSTP is invalid, so is there another way that leads to a different outcome?

There is another way, and an elegant one too, but it depends on getting Bayes theorem into another form. Readers who are not interested in the derivation of Equation 19 can skip directly to it. I will use Equation 19 to solve the problem of the two physicians [14, 17].

It is convenient to start from Equation 1 rather than Equation 2. Consequently, I rewrite Equation 1 below as Equation 15 but this time using  $H_1$  to denote the hypothesis proposed by Physician 1. I also use  $H_2$  to denote the



Hypothesis 2 is 2.49 times as likely to be true than is Hypothesis 1. Thus, although the NHSTP settles the argument in favor of Physician 1, Equation 19 settles the argument in favor of Physician 2.

Of course, we might not be satisfied with 20 trials. Suppose our two physicians want to have more confidence, so they perform the study again. Suppose they again cure 13 patients out of 20 attempts. What is the new posterior ratio?

The posterior ratio from the previous experiment becomes the new prior ratio for the new experiment so that the updated prior ratio is now 2.49. Multiplying by the data ratio, which also is 2.49, renders an updated posterior ratio of 6.20. In other words, Hypothesis 2 is 6.20 times more likely to be correct than is Hypothesis 1. This process can go on as long as our two physicians have the patience and grant money to continue to perform new experiments. Equation 21 summarizes the process of continually updating after each experiment where  $k$  is the number of experiments [18].

$$\frac{P(H_2|F)}{P(H_1|F)} = \frac{P(H_2)}{P(H_1)} \prod_i^k \frac{P(F_i|H_2)}{P(F_i|H_1)} \quad (21)$$

As an example of Equation 21, suppose that our physicians perform a total of 5 experiments and the data ratios from these are 2.49, 2.49, 4.23, 2.11, and 2.78, respectively. Commencing at a prior ratio of 1 to provide a fair playing field for the two hypotheses, the final posterior probability is  $1 \times 2.49 \times 2.49 \times 4.23 \times 2.11 \times 2.78 = 153.84$ . In other words, the data from the five experiments indicate that Hypothesis 2 is more likely in the extreme than is Hypothesis 1. The case of the two physicians provides a nice demonstration of the advantages of Bayesian reasoning as a way to test competing medical hypotheses with updating as data continue to be collected.

## V. AUXILIARY BAYESIAN ISSUES IN MEDICINE

Although the foregoing sections demonstrate that Bayesian approaches can be useful in medicine, there are auxiliary issues and complications. The goal of the present section is to be up front about those issues and complications.

### What is the Relevant Base Rate?

In the foregoing Bayesian example where the probability of the HNP disorder was less than 8% given a positive MRI finding, I used the population prevalence as the base rate. Arguably, there are other base rates that could have been used. For example, perhaps the base rate of HNP disorder is greater among patients who seek treatment for back pain than among people in the general population. For example, suppose we quadruple the .03 value used earlier and imagine that the augmented base rate is actually .12. How much does the base rate matter?

In fact, it matters a lot. If we again plug the relevant values obtained from the literature into Equation 2, with the exception of using .12 instead of .03 for the base rate, the final probability that the patient has HNP increases to .268 (see below).

$$P(D|T) = \frac{(.120)(.804)}{(.120)(.804) + (.300)(1 - .120)} = .268$$

There are three interesting points that this example suggests. First, of course, the value of .268 is much larger than the previously obtained value of .077; in fact it is 3.48 times larger. Base rates really matter! Second, despite quadrupling the base rate and substantially increasing the posterior probability of the disorder given the positive test result, the probability of .268 is nevertheless a low number. Consider, as often is so, that a physician uses the positive MRI finding to recommend surgery. Although the physician will benefit substantially through remuneration for the surgery, there is a strong chance of complications for the patient. Few people would argue seriously that it is worthwhile for the patient to take the substantial risks of surgery when there is a 73.2% (100% - .268%) chance that the patient does not even have the disorder.

Suppose that the base rate that the researcher or doctor wishes to use is not known. In that case, it is not clear what value to plug into Equation 2 and so the researcher or doctor would seem to be sunk. However, even in this unenviable position, Bayesian thinking still has the possibility to be useful. The reason is that the researcher or physician can plug in a variety of values just to see what happens. Again using the HNP example, consider the following cases where we use values of .1, .2, .3, .4, .5, .6, .7, .8, and .9, respectively for  $P(D)$ .

$$P(D|T) = \frac{(.1)(.804)}{(.1)(.804) + (.300)(1 - .1)} = .229$$

$$P(D|T) = \frac{(.2)(.804)}{(.2)(.804) + (.300)(1 - .2)} = .401$$

$$P(D|T) = \frac{(.3)(.804)}{(.3)(.804) + (.300)(1 - .3)} = .535$$

$$P(D|T) = \frac{(.4)(.804)}{(.4)(.804) + (.300)(1 - .4)} = .641$$

$$P(D|T) = \frac{(.5)(.804)}{(.5)(.804) + (.300)(1 - .5)} = .728$$

$$P(D|T) = \frac{(.6)(.804)}{(.6)(.804) + (.300)(1 - .6)} = .801$$

$$P(D|T) = \frac{(.7)(.804)}{(.7)(.804) + (.300)(1 - .7)} = .862$$

$$P(D|T) = \frac{(.8)(.804)}{(.8)(.804) + (.300)(1 - .8)} = .915$$

$$P(D|T) = \frac{(.9)(.804)}{(.9)(.804) + (.300)(1 - .9)} = .960$$

The example brings out certain points. In general, and consistent with Figure 1, as the base rate for HNP increases, so does the posterior probability of HNP. But there are more specific issues too. Remember that when we used the base rate that actually was obtained from the literature (3%), the posterior probability of HNP was less than 8%. If we bump the base rate up to 10%, the posterior probability rises to 22.9%. Nevertheless, even this enhanced value fails to provide a strong case for diagnosing HNP from a positive MRI test. Even if we bump up the base rate to 20% (almost seven-fold the base rate in the literature), the probability of HNP given a positive MRI test is 40.1%, which is still well under 50%. Does it make sense to draw a strong conclusion about the presence of HNP based on a method that works less well than a coin toss? The larger point here is that even if one has reason to suspect that a base rate obtained in the literature is an underestimate, possibly because the reference group is not the ideal reference group, Bayesian reasoning can still be employed. The researcher can plug enhanced base rates into the theorem to determine whether they cause the estimated posterior probability of the disorder to rise to a level that is acceptable for drawing a strong conclusion. If plugging in the largest “reasonable” base rate still fails to bring the estimated posterior probability of the disorder to an acceptable level, the researcher or physician has a reason not to base diagnoses on positive test results for the disorder of concern.

### **What do we mean by “Probability?”**

A potential criticism of applying Bayesian reasoning to determining the probabilities of null hypotheses is philosophical. The idea stems from an assumption that a hypothesis can be correct or incorrect but it cannot have a probability. If hypotheses cannot have probabilities, then how is it possible to make a Bayesian criticism of null hypothesis significance tests? In addition, how is it possible to justify using Equation 19 or Equation 21 to estimate probabilities of hypotheses given findings? Although there is insufficient space for a full discussion of these philosophically and historically complex issues [14], I hope to provide sufficient information for the reader to have his or own thoughts about the issue. In short, much depends upon what one means by the term, “probability.”

To commence, consider Einstein’s famous quotation that “God does not play dice with the universe.” Einstein was not claiming to know everything but he was claiming that if a being did know everything, and had infinite information processing power, that being would be able to predict the future with certainty. Thus, there is no randomness in the universe and any kind of statement indicating uncertainty is really a claim about lack of knowledge. Given perfect knowledge, the probability of anything would be zero or one but people do not have perfect knowledge, and so they use probabilities between zero and one to indicate the imperfect state of their knowledge. To sum up in a word, Einstein’s famous quotation suggests that a probability statement is really an “epistemic estimate;” that is, a probability statement is an estimate of one’s state of knowledge (or lack thereof). From Einstein’s perspective, then, it makes perfect sense to talk about probabilities of hypotheses because we have imperfect knowledge about whether they are true, though obtained data can suggest that it is more or less likely that they are true than was thought before the data were collected. Sometimes, the view of probability as pertaining to people’s subjective mental states is termed the “belief state” view of probability.

On the other hand, few modern physicists would agree with Einstein on this matter. There is much evidence that randomness really does exist in the universe. Even if a being knew everything there was to know and had infinite processing capacity, that being nevertheless would be incapable of predicting the future perfectly. Randomness in the universe would render the being's predictions to be wrong some of the time. From the point of view of modern physics, then, there really is randomness in the universe that can be distinguished from lack of knowledge on the part of the researcher. If one believes that there really is randomness in the universe, then probabilities are not statements about one's degree of knowledge or lack thereof. Rather a statement about a probability really does reflect that there is some randomness in the universe. When a physicist states that an electron has a particular probability of being in a certain orbital, this is not an epistemic estimate due to lack of knowledge but indicates that even if the physicist knew everything there was to know about the universe, the electron nevertheless might or might not be in the specified orbital.

There are at least two ways to conceptualize probability in a way that asserts that randomness is not just within the researcher's mind but actually is "out there" [19]. As opposed to defining a probability as indicating a belief state, some philosophers would claim that a probability is a relative frequency. For example, if there are 10 white balls and 30 black balls in a bowl, the probability of drawing a white ball is simply the relative frequency of white balls to total balls; i.e.,  $P(\text{White Ball}) = \frac{10}{10+30} = .25$ .

Others would prefer to describe probabilities as "propensities." The idea of conceptualizing a probability as a propensity is that events might not have happened yet and so one is considering how a system generates events—the system's propensity to cause different classes of events to occur [19]. For example, consider a fair coin. From a probability-as-propensity point of view, it does not make sense to talk about frequencies of heads versus tails because they have not happened yet. But it does make sense to say that a fair coin has the propensity to produce equal frequencies of heads or tails when properly flipped.

In the event that a researcher insists that a hypothesis is either true or false but cannot have a probability, does this render a Bayesian approach as invalid? To approach an answer to this point, let me first point out that even if one disallows a Bayesian approach to hypothesis testing, such disallowing fails to rescue the traditional NHSTP. That is, as I explained earlier, there remains no logical basis for using  $p$ -values to reject null hypotheses. But such disallowing also is a problem for what might be thought of as a philosophically naïve Bayesian approach that cannot account for cases where belief states and real probabilities are both relevant because of the insistence that everything that goes into the Bayesian calculus must be a probability of some type. But it is possible to propose a more flexible Bayesian approach that allows for a mixture of epistemic estimates and real probabilities [14]. To see that this is so, let us revisit the debate between the two physicians who disagreed about the probability that a particular medical procedure would cure a particular disorder. In the example, I derived and used Equation 19 to settle the debate. However, I arguably committed a philosophical crime by setting the prior ratio at 1 to give each hypothesis a fair chance at winning out over the other one. Arguably, each hypothesis is either true or false and does not have a probability, so I had no business assigning values. How can this potential criticism be addressed?

One way is to just admit that the prior ratio is not based on probabilities but rather on epistemic estimates [14, 17]. In that case, we would rewrite Equation 19 but using  $E$  for "epistemic estimate" instead of  $P$  for "probability," in the appropriate places. This renders Equation 22 below.

$$\frac{P(H_2|F)}{P(H_1|F)} = \frac{P(H_2) P(F|H_2)}{P(H_1) P(F|H_1)} \rightarrow \frac{E(H_2|F)}{E(H_1|F)} = \frac{E(H_2) P(F|H_2)}{E(H_1) P(F|H_1)} \quad (22)$$

Equation 21 is of much philosophical interest. For those who are not overly concerned with exactly the meaning of "probability," it likely is easier to use Equation 19 than to use Equation 22. But for those with a philosophical bent, Equation 22 brings to light important issues. Notice that the base rate probabilities of the two hypotheses are replaced with epistemic estimates of the two hypotheses. Likewise, the final result is not a ratio of posterior probabilities but is rather a ratio of posterior epistemic estimates. The product of an epistemic estimate and a probability is an epistemic estimate and not necessarily a probability. For those who do not distinguish between epistemic estimates and probabilities, Equation 22 is a meaningless exercise but for those who do, Equation 22 can be used to justify the Bayesian calculus even when epistemic estimates and probabilities are mixed. But there is a cost. For example, in this latter case, there is at least one incorrect interpretation of the experiment where there were 13 successes out of 20 trials. It would be incorrect to say, "The probability that the hypothesis suggested by Physician 2 is correct is 2.49 times greater than is the probability that the hypothesis suggested by Physician 1 is to be correct." But it would be correct to say, "The experimental evidence is 2.49 times stronger in favor of the hypothesis suggested by Physician 2 than in favor of the hypothesis suggested by Physician 1." Alternatively, it also

would be correct to say, “The epistemic status of the hypothesis suggested by Physician 2 is 2.49 times as great as the hypothesis suggested by Physician 1.”

## VI. CONCLUSION

I have argued that Bayesian analysis can help medical researchers and physicians in two broad respects. First, there is a wealth of data pertaining to base rates and false negative and positive test results that has not been properly exploited to the benefit of medical science. As in the HNP example, important advances can be made merely by finding the relevant data and using Bayes’ theorem to calculate the probability of the disorder given a positive test result. Second, like many sciences, medicine depends entirely too much on the NHSTP that is blatantly illogical. A Bayesian perspective illuminates precisely where the logical flaw lies. The NHSTP has done incalculable damage to the field and medical researchers ought to stop using it. There are Bayesian alternatives (e.g., Equation 19) that are preferable and that often can be used. Finally, although the point of this article, and the Special Issue more generally, is to promote Bayesian reasoning in medicine, nothing is perfect in every way and this is so even with Bayesian reasoning. I addressed some special problems and suggested ways of addressing them.

## REFERENCES

- [1] S. D. Boden, D. O. Davis, T. S. Dina, N. J. Patronas, and S. W. Wiesel, “Lumbar spine in asymptomatic subjects: A prospective investigation,” *J. Bone Joint Surg Am*, vol. 72, no. 3, pp. 403-407, 1990.
- [2] M. C. Jensen, M. N. Brant-Zawatzki, N. Obuchowski, M. T. Modic, D. Malkasian, and J. S. Ross, “Magnetic resonance imaging of the lumbar spine in people without back pain,” *N Engl J Med*, vol. 331, no. 2, pp. 69-73, 1994.
- [3] G. B. Andersson, “Epidemiology of spinal disorders,” in J. Frymoyer, ed., *The Adult Spine: Principles and Practice*, pp. 107-146, Raven Press, New York, 1991.
- [4] R. C. Lawrence, D. T. Feseon, C. G. Helmick, L. M. Arnold, H. Choi, R. A. Deyo, S. Gabriel, R. Hirsch, M. C. Hochberg, G. G. Hunder, J. M. Jordan, J. N. Katz, H. M. Kremers, and F. Wolfe, “Estimates of the prevalence of arthritis and other rheumatic conditions in the United States Part II,” *Arthritis & Rheumatism*, vol. 58, no. 1, pp. 26-35, 2008.
- [5] J. W. Frymoyer, “Epidemiology,” in J. W. Frymoyer, and S. L. Gordon, eds, *New Perspectives on Low Back Pain*, Workshop, Airlie, Virginia, May, pp. 19-33, 1988 (Symposium).
- [6] D. G. Borenstein, D. G. Wiesel, and S. D. Boden, “*Low Back Pain-Medical Diagnosis and Comprehensive Management 2<sup>nd</sup> Ed.*,” WB Saunders Co., Philadelphia, 1995.
- [7] N. Boos, R. Rieder, S. Volker, K. F. Spratt, N. Semmer and M. Aebi, “The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations,” *Spine*, vol. 20, no. 24, pp. 2613-2625, 1995.
- [8] D. Trafimow, and J. H. Trafimow, “The shocking implications of Bayes’ theorem for diagnosing herniated nucleus pulposus based on MRI scans,” Submitted manuscript.
- [9] D. Trafimow, “Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes’s theorem,” *Psychological Review*, vol 110, no. 3, pp. 526-535, 2003.
- [10] R. A. Fisher, *Statistical methods and scientific inference* (3<sup>rd</sup> ed), London, Collier Macmillan, 1973.
- [11] D. Bakan, 1966. “The test of significance in psychological research,” *Psychological Bulletin*, vol. 66, no. 6, pp. 423-437, 1966.
- [12] J. Cohen, “The earth is round ( $p < .05$ ),” *American Psychologist*, vol. 49, no. 12, pp. 997-1003, 1994.
- [13] T. Sellke, M. J. Bayarri, and J. O. Berger, “Calibration of  $p$  values for testing precise null hypotheses,” *American Statistician*, vol. 55, no. 1, pp. 62-71, 2001.
- [14] D. Trafimow, “Using epistemic ratios to evaluate hypotheses: An imprecision penalty for imprecise hypotheses,” *Genetic, Social, and General Psychology Monographs*, vol 132, no.4, pp. 431-462, 2006.
- [15] J. Krueger, “Null hypothesis significance testing: On the survival of a flawed method,” *American Psychologist*, vol 56, no. 1, pp. 16-26, 2001.
- [16] D. Trafimow, and S. Rice, “A test of the null hypothesis significance testing procedure correlation argument,” *Journal of General Psychology*, vol. 136, no. 3, pp. 261-269, 2009.
- [17] D. Trafimow, “The ubiquitous Laplacian assumption: Reply to Lee and Wagenmakers,” *Psychological Review*, vol. 112, no. 3, pp. 669-674, 2005.
- [18] D. Trafimow, D. “On making assumptions about auxiliary assumptions: Reply to Wallach and Wallach,” *Theory and Psychology*, vol. 20, no. 5, pp. 707-711, 2010.
- [19] K. R. Popper, *Realism and the aim of science*, London, Routledge, 1983.