All tests are imperfect: Proper interpretation makes some useful

Song S. Qian[a,1], **Jeanine M. Refsnider**[a], **Jennifer A. Moore**[b], **Gunnar R. Kramer**[a], **and Henry M. Streby**[a]

[a]Department of Environmental Sciences, The University of Toledo, 2801 W. Bancroft Street, Toledo, OH 43606, USA; [b]Department of Biology, Grand Valley State University, 3300a Kindschi Hall of Science, Allendale, MI 49401, USA

This manuscript was compiled on January 7, 2018

Tests with binary outcomes (e.g., positive versus negative) to indicate a binary state of nature (e.g., disease agent present versus absent) are common. These tests are rarely perfect: the chance of a false positive (negative) is always larger than 0. Interpreting an imperfect result requires proper accounting for uncertainty using the conditional probability formula, the Bayes' rule. When the objective is to understand the status of the test subject, we use the Bayes' rule to estimate the probability of, for example, disease agent present given the test result is positive. A positive result does not always suggest that the disease agent is more likely to be present; not only the two error rates but also the prevalence of the disease in the relevant population will affect the meaning of a positive (negative) result. When the objective is to understand the status of the population represented by the test subjects (e.g., prevalence of the disease) the more general version of the Bayes' rule is used to quantify the population characteristics. Using numerical examples, we illustrate the post-test steps necessary for making the imperfect test results meaningful.

Bayes' rule | conditional probability | false negative | false positive | uncertainty

In both scientific research and routine daily decision-making, we depend on results of various tests. Tests come in all forms and shapes. For example, doctors may test a patient's blood for the presence of a disease marker, environmental engineers may test water at a swimming beach for the presence of fecal coliform bacteria as an indicator of sewage pollution, ecologists may survey for invasive species, a geologist may drill a test well exploring for oil, a pollster may take an opinion poll to evaluate the viability of a political candidate, and so on. A unifying feature of these tests is that they are imperfect: the test result is likely correct but not always. In most cases, we can simplify a test result to be either positive or negative. For example, a positive blood test indicates the existence of the disease marker and a positive opinion poll result indicates that the candidate is likely to win (with more than 50% popular support). A test result is imperfect because a test is always based on a sample (e.g., a blood sample, a water sample, or a sample of 1,000 potential voters). A sample can misrepresent the population. A water sample from a polluted water source may contain no fecal coliform bacteria by chance, thereby leading to a false negative result. Likewise, a water sample from a clean water source may be contaminated unintentionally by researchers during the sampling process or in the lab. The subsequent positive result is then classified as a false positive (i.e., the water source is incorrectly classified as "polluted"). Almost all tests used in scientific research are imperfect and, therefore, such sampling and experimental errors are unavoidable. Given this imperfection of tests, we face a challenge when interpreting the test result: how can we use a potentially incorrect result to draw inference or make a decision? Is a positive fecal coliform test result truly indicating that the water is polluted by domestic sewage, or is it a false positive? How we interpret the result will affect how we decide, for example, when to issue a public health warning for an affected recreational beach.

The imperfection of a test is routinely quantified with two error rates: the rate of a false positive and the rate of a false negative. Scientists in all fields recognize the imperfection, and used various terms to describe the two types of errors. For example, in the early days of World War II, the US Army developed the receiver operating characteristic (ROC) curve to identify the optimal threshold for determining whether a radar signal was from a Japanese aircraft. The two axes of the ROC curve are errors of omission (false negative) and commission (false positive). For simplifying the discussion, we will define the following terms. First, we use present or absent to represent the state of the world we are trying to infer: a present indicates the presence of an agent of interest (e.g., fecal coliform, more than 50% popular support, an enemy aircraft, and so on) and an absent means the absence of the agent. When conducting a test, the test result is either positive (indicating the state of the world is present) or negative (absent). A false positive rate tells us how likely it is that a test would lead to a positive result when the state of the world is absent and a false negative rate is the likelihood of a negative result when the state of the world is present. When a test is carried out, we want to use the test result (either positive or negative) to infer the state of the world. The imperfection of the test leads to uncertainty in the subsequent interpretation and inference.

Significance Statement

All tests are imperfect and the associated uncertainty is manifested in probabilities of false positives and false negatives. Results of an imperfect test are meaningful only when these two error rates, as well as our knowledge of the subject, are properly accounted. We discuss statistical methods for incorporating this information under three scenarios. The results are applicable to almost all tests. The importance of properly interpreting results from imperfect tests is universal, although how to handle the uncertainty is inevitably case-specific. The statistical considerations not only will change the way we interpret test results, but also how we plan and carry out tests that are known to be imperfect.

Author contributions: Qian developed the idea based on the work of Moore. The first draft was written after lengthy discussions among Qian, Refsnider, and Moore. Kramer and Streby commented on the first draft of the paper and contributed substantially in the writing of the final draft. All authors participated in discussions on the presentation of the work.

The authors declair no conflicts of interest in this work.

[1]To whom correspondence should be addressed. E-mail: song.qianutoledo.edu

Conditional Probability and the Bayes' Rule

Properly handling the uncertainty of the test result is the realm of probability and statistics. We use probability to quantify the uncertainty and rules of probability allow us to make inferences. Using the probability language to describe an imperfect test, the rate of a false positive is the probability of a positive test result when the underlying state of the world is absent. Likewise, the rate of a false negative is the probability of a negative test result when the state of the world is present. These two probabilities are examples of a *conditional* probability. To summarize the rules of conditional probabilities, we use "p" to represent the state of the world being present, "a" to represent absent (Figure 1(i)), "+" to represent a positive test result, and "−" a negative result (Figure 1(ii)). A false positive probability is symbolized as $\Pr(+|a)$, and a false negative probability is $\Pr(-|p)$. These two conditional probabilities characterize the quality of the test.

When a test is carried out, we observe either a "+" or a "−". What we want to know is how likely the state of the world is p when observing a + and how likely the state of the world is a when observing a "−". These are also conditional probabilities: $\Pr(p|+)$ and $\Pr(a|-)$. These two conditional probabilities are the basis for interpretation and inference of imperfect tests.

The Bayes' rule (1) is the probability rule connecting these two groups of conditional probabilities (Figure 1(iii)):

$$\Pr(p|+) = \frac{\Pr(p)\Pr(+|p)}{\Pr(p)\Pr(+|p) + \Pr(a)\Pr(+|a)} \quad [1]$$

and

$$\Pr(a|-) = \frac{\Pr(a)\Pr(-|a)}{\Pr(a)\Pr(-|a) + \Pr(p)\Pr(-|p)} \quad [2]$$

We will focus on equation Eq. (1). The test result is either positive or negative; consequently, $\Pr(+|p) = 1 - \Pr(-|p)$. In addition, $\Pr(a) = 1 - \Pr(p)$. The Bayes' rule (eq. Eq. (1)) suggests that in addition to the false positive and false negative probabilities we must also know $\Pr(p)$ in order to calculate $\Pr(p|+)$. In statistics, $\Pr(p)$ is a marginal probability – the probability of the state of the world being present regardless of the test result (or before we carried out the test). This probability can be interpreted as, for example, the prevalence of a disease in a population or our uncertainty with regard to the state of the world before a test is carried out. For example, when testing snakes for a snake fungal disease (*Materials and Methods*), we can interpret $\Pr(p)$ as the prevalence of the disease in the population. The Bayes' rule suggests that $\Pr(p)$ is necessary when interpreting the test result (see SI-A.).

The Prior Probability and Statistical Inference

A point of contention in using the Bayes' rule is the meaning of the prior probability $\Pr(p)$. In SI-A., we interpreted the prior as the fraction of individual snakes in the population that are infected with the fungal disease. When this fraction changes, the proportion of true positives (positives from infected individuals) also changes. When the prior has a clear physical meaning and can be measured, the use of the Bayes' rule is widely accepted (2). When the prior is difficult to estimate or the physical meaning is ambiguous, the use of prior used to be controversial. Increasingly, we recognize that estimating the prior is a means for proper use of relevant information in an analysis (SI-B.). Regardless of the meaning of the prior, the Bayes' rule highlights the need of quantifying the prior $\Pr(p)$ in order to properly interpret a test result; whether we call the quantity a prior, a marginal probability, or the prevalence is irrelevant. The proper interpretation of the test result and the use of the test result for inference requires a proper statistical treatment. As in all statistical applications, the first step is to represent the scientific hypothesis using a statistical model (with parameters). The proposed statistical model, in turn, will decide how we use data to estimate model parameters and how the model can be verified.

The Purpose of A Test

Understanding and defining the objective of a test is the key to deciding what statistical model to use. When analyzing results from an imperfect test, there are at least two different objectives: (1) testing for an individual subject, that is, to estimate $\Pr(p|+)$ (e.g., whether a patient has a particular disease agent), and (2) testing to learn about a population, that is, to estimate $\Pr(p)$ (e.g., the prevalence of SFD in the rattlesnake population in Michigan).

Tests for Individual Subjects. When a doctor tests a patient for a disease, the objective is to determine the likelihood that the patient has the disease. With known characteristics of the test (i.e., $\Pr(+|a)$ and $\Pr(-|p)$) and the prevalence of the disease in the population, we can use the Bayes' rule to calculate the conditional probability $\Pr(p|+)$ for a positive result. In this situation, we want a test that will result in a $\Pr(p|+)$ larger than 0.5. That is, a positive result should suggest that the patient is more likely to have the disease than not. Another way to express this condition is that the odds ratio should be larger than 1: $\frac{\Pr(p|+}{1-\Pr(p|+)} > 1$. More generally, we can require that $\frac{\Pr(p|+}{1-\Pr(p|+)} > \gamma$ before we consider the test to be useful (e.g., prescribe treatment upon a positive result). Using the Bayes' rule, we can express this requirement in terms of the rates of false positive and false negative, as well as the prevalence.

$$\Pr(+|a) < \frac{1-\gamma}{\gamma}\frac{\Pr(p)}{1-\Pr(p)}(1 - \Pr(-|p)) \quad [3]$$

In other words, the inequality set by equation Eq. (3) must be met before a test is considered useful (Figure 2).

Tests for Estimating Population Parameter(s). In many cases, we test multiple individuals of a population in order to understand the characteristics of the population. In the snake example (*Materials and Methods*), we are interested in estimating the prevalence of the fungal disease in the population. That is, we want to estimate the prior $\theta = \Pr(p)$. There are two approaches to estimate θ. One is using inverse Bayes formulae (IBF)(3), which is a simple and straightforward method for situations where the true state of the world of the sampled individuals (e.g., whether a snake is truly infected with SFD) can be ascertained (perhaps with additional experimentation). Specifically, once we know the true state of the world of the samples, we can use the test results to estimate the posterior directly from the data (SI-C.). That is, if we have

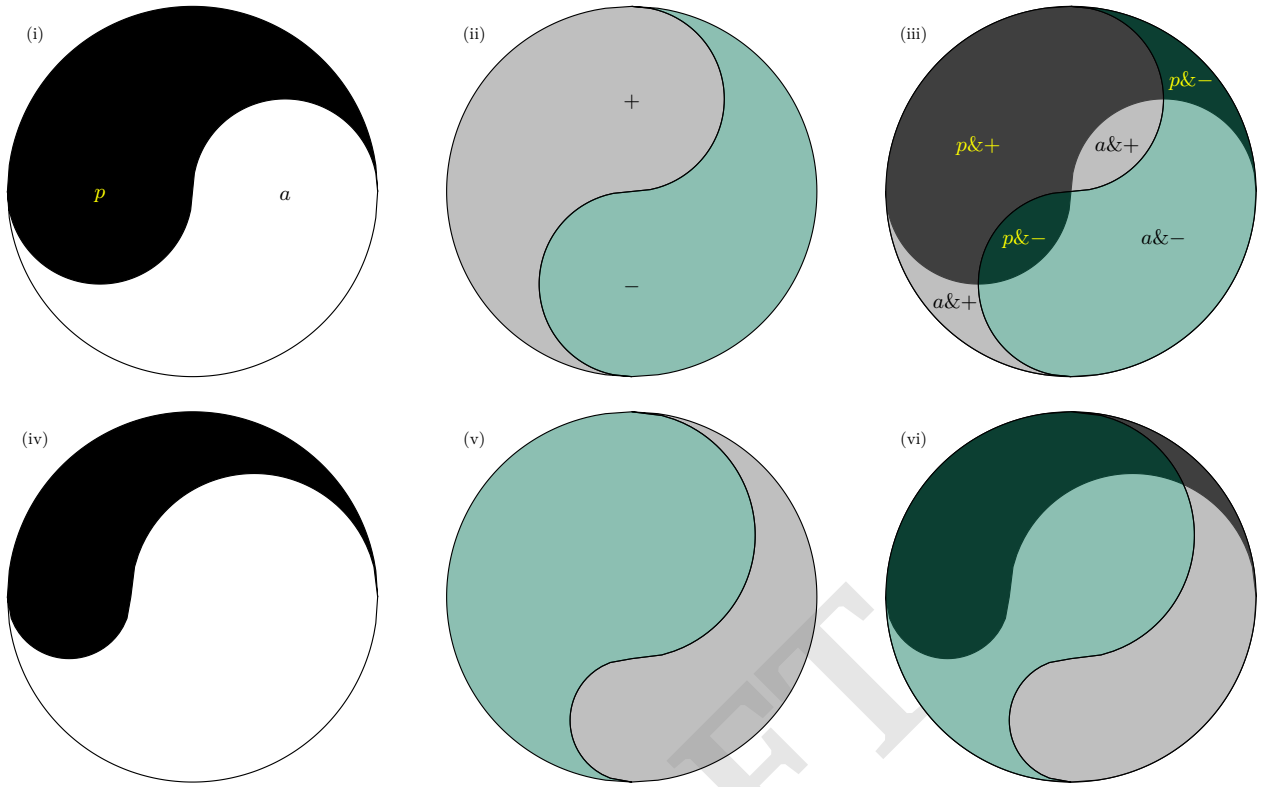

Fig. 1. A graphical depiction of an imperfect test. The real state of the world is either p (present) or a (absent) (i), and the test result is either $+$ or $-$ (ii). The imperfection of the test makes the interpretation of a test result contingent on information regarding the accuracy of the test, as well as the relative size of p and a (iv-vi).

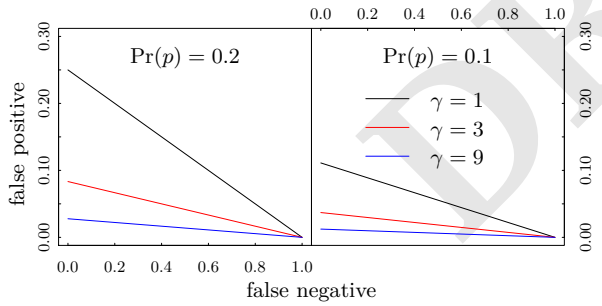

Fig. 2. A graphical representation of Eq. (3)

$\Pr(p|+), \Pr(+|a)$, and $\Pr(-|p)$, we can calculate $\Pr(p)$ using IBF.

When the true state of the world cannot be ascertained, we need to approach the problem from a different angle. The objective of the study is now the estimation of $\theta = \Pr(p)$, a continuous variable, based on observed positives and negatives. In statistical terms, we tested n snakes and observed y positives, from which we wish to estimate the prevalence θ. We start the process by proposing a statistical model describing the data generating process. In the case of analyzing test results, the data are the number of positive results from a total number of subjects. The statistical model describing the distribution of the data is the binomial distribution. The model is parameterized by a single parameter – the probability of observing a positive result. The quantity of interest is the probability of infection. How the parameter of interest and the binomial model parameter are linked depends on what we know (SI-D.). If the test is perfect, that is, we know the rates of false positive and false negative are both 0, we have a simple binomial-beta model and the parameter can be easily estimated. This model is often the first model in an introductory Bayesian statistics book (4). When the complexity of the data generation process increases, the simple model needs to be modified. If the test is imperfect and rates of false positive and false negative are known, the posterior distribution of θ cannot be represented by a commonly seen probability distribution. But the posterior distribution can be numerically evaluated and graphed for inference (SI-D. and E.). For example, suppose we tested 20 snakes for fungal disease and 5 were positive. If the test has a false negative rate of 5% and a false positive rate of 2%, and our initial guess of the prevalence is 10% based on a previous study of 20 snakes (our prior of the prevalence is a beta distribution with parameters $\alpha = 2$ and $\beta = 18$), the posterior distribution is numerically estimated and shown in Figure 3.

When one of the two error rates is unknown and needs to be estimated, we must estimate a two-dimensional joint probability distribution. Computation is more intense, although we can still graphically display the joint distribution. If we are uncertain about the false positive probability in the case shown in Figure 3, we can use a beta distribution to describe the uncertainty. For example, we may use a $beta(2, 100)$ to represent a false positive rate with mean about 0.02 and a stan-

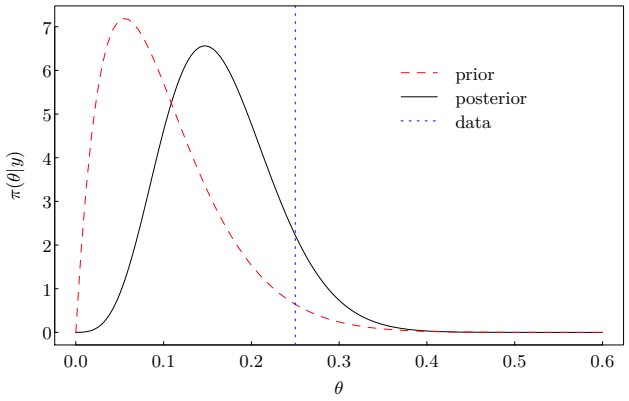

Fig. 3. Numerically estimated probability density function of the posterior distribution of the population prevalence/prior (θ)

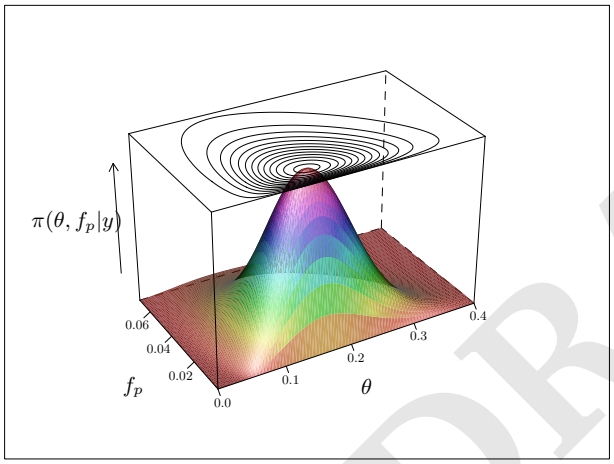

Fig. 4. A perspective plot of the numerically estimated joint distribution of the false positive probability (f_p) and population prevalence/prior (θ)

dard ard deviation of 0.014 (a middle 95% range of 0.0024 and 0.054). The numerically estimated joint posterior distribution is shown in Figure 4.

When both rates are unknown, the posterior distribution becomes a three-dimensional joint distribution, and we must resort to a modern Bayesian computation method to estimate the parameters and a numerical summary of these parameters to understand the distribution (SI-E.) .

The steps of moving from the simplest binomial-beta model to the model simultaneously estimating three parameters are typical in statistical learning. These steps should also be taken in applying statistics. When using imperfect tests, we need to understand why and how the test is imperfect and derive the appropriate model accordingly. Such a process is often tedious and iterative, a message we almost always miss when teaching or taking a statistics course.

Practical Implications

Although our discussions were made with specific examples, the problem of an imperfect test is universal. How we interpret results from an imperfect test depends on how the data were collected and for what purpose. When individual test subjects are of concern, the uncertainty should be presented in terms of a conditional probability (i.e., $\Pr(p|+)$). The quality of the test is the key for proper interpretation of the test result. This is not just a problem of medical diagnostic tests. The statistical principle is the same in any situation where results from an imperfect test are applied to a specific subject. When an imperfect test is used to infer a population parameter such as the prevalence of snake fungal disease in rattlesnake populations in Michigan, test results are raw data to be further processed to estimate the parameter of interest. The uncertainty associated with the imperfect test is represented by the posterior probability distribution. Depending on the nature of the test and the ease of determining the true status of a test subject, we have different computational needs. In other words, the proper use of an imperfect test requires us to fully consider all available information and properly structure the statistical analysis based on the objective of the study, just as the interpretation of the p-value from a null hypothesis test with low power (5) should be case-specific.

Summary – A Guide to Practitioners

Our discussion does not change how a practitioner conducts a test. Rather, we argued that the test result should be properly interpreted based on the quality of the test procedure, the objective of the test, and the state-of-the art understanding of the subject matter, and report the result accordingly (perhaps to avoid headline-catching, yet erroneous, statements).

Test Target	Individual subject	Population
Test objective	$\Pr(p\|+)$ Does the subject have the disease?	$\Pr(p)$ The disease prevalence in the population
Test result	+/-	+/- (multiple subjects)
Knowledge needed to interpret test results	$\Pr(p)$, quality of the test $\Pr(+\|a), \Pr(-\|p)$	$\Pr(+\|a), \Pr(-\|p)$, and an educated guess of $\Pr(p)$
Additional steps (beyond test results)	Bayes' rule	IBF if subject's true state can be ascertained, or Bayesian computation when true state is infeasible
How can the test be improved?	improved testing method	... and a better guess of $\Pr(p)$

Materials and Methods. We used a study of fungal disease in a Michigan rattlesnake population as an example. Snake fungal disease (SFD), caused by the fungus *Ophidiomyces ophiodiicola*, is an emergent pathogen known to affect at least 30 snake species from 6 families in eastern North America and Europe (6, 7). SFD was detected in eastern massasaugas (*Sistrurus catenatus*), a small, federally-threatened rattlesnake species, in Michigan in 2013 (8). The estimated SFD prevalence ranges from 3-17% in three Michigan populations (9).

A commonly used method for detecting SFD is quantitative PCR (qPCR) to detect the fungal DNA using a skin swab. The method often leads to a false negative because swabbing can miss the disease agent. Hileman et al.(9) show that a

single swab of an eastern massasauga with clinical signs of SFD (skin lesions) can result in a false negative result 73% of the time; a positive result (detecting fungal DNA on an individual snake) does not always indicate that the individual has SFD.

For the purpose of discussion, we used a sample of 20 snakes, 5 which tested positive for SFD. As the effectiveness of using qPCR for testing SFD is still under study, we used optimistic hypothetical rates of false positive (2%) and false negative (5%).

Statistical methods are fully discussed in the Supporting Information, including:

A. A numerical explanation of the Bayes' rule,

B. An intuitive explanation of the Bayesian prior,

C. Inverse Bayes Formulae (IBF) for estimating the population parameter (the prior),

D. An exposition of Bayesian inference, and

E. Bayesian computation.

Computer code with detailed annotation can be found at GitHub.

ACKNOWLEDGMENTS. We thank Ian Clifton, Jessica Garcia, and Tyara Vasquez for insightful discussions.

1. DeGroot M (1986) *Probability and Statistics*. (Addison-Wesley Publishing Company, Inc.), Second edition, p. 723.
2. Efron B, Morris C (1977) Stein's paradox in statistics. *Scientific American* 236:119–127.
3. Tan M, Tian G, Ng K (2010) *Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation*. (Chapman and Hall/CRC Press, Boca Raton, FL, USA).
4. McElreath R (2016) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. (Chapman & Hall/CRC Press).
5. Halsey L, Curran-Everett G, Vowler S, Drummond G (2015) The fickle P value generates irreproducible results. *Nature Methods* 12:179–185.
6. Allender M, Raudabaugh D, Gleason F, Miller A (2015) The natural history, ecology, and epidemiology of *Ophidiomyces ophiodiicola* and its potential impact on free-ranging snake populations. *Fungal Ecology* 17:187–196.
7. Lorch J, et al. (2016) Snake fungal disease: an emerging threat to wild snakes. *Philosophical Transactions of the Royal Society B* 371(1709):20150457.
8. Tetzlaff SJ, Allender M, Ravesi M, Smith J, Kingsbury B (2015) First report of snake fungal disease from Michigan, USA involving massasaugas, *Sistrurus catenatus* (Rafinesque 1818). *Herpetology Notes* 8:31–33.
9. Hileman E, et al. (2017) Estimation of *Ophidiomyces* prevalence to evaluate snake fungal disease risk. *The Journal of Wildlife Management* p. doi:10.1002/jwmg.21345.

Supporting Information (SI Appendix)

A. A simple explanation of the Bayes' rule

- The Bayes' rule is a rule about conditional probability. We use a simple numeric example to explain equation Eq. (1). In our example of using the qPCR method to detect SFD in a snake, the test is imperfect. Suppose that the false positive rate is 5% and the false negative rate is 3% (much improved over the current test). Suppose we also know, based on similar studies on the same snake species in the Midwest, the prevalence of the disease is about 4%. In order to interpret a positive or negative test result, we use the Bayes' rule. We can explain the Bayes' rule as a straightforward account of the expected number of true positives and false positives. Assuming there are 10,000 snakes in this population, our prior knowledge suggests that 400 snakes have the disease and 9,600 snakes do not. (We used an unrealistically large population number to avoid non-integer numbers.) With a false negative probability of 3%, we expect to have 12 false negatives and 388 true positives. Likewise, among the 9,600 healthy snakes, a 5% false positive rate will result in 480 false positives. If we test all 10,000 snakes, we would expect 868 positives, and only 388 of them are truly infected. For a randomly caught snake, a positive result would suggest that the chance that the snake is truly infected is 388/(388+480) or 0.447. In other words, a snake with a positive test result is (slightly) less likely to be infected with the disease than not infected. The Bayes' rule leads to the same result. The numerator of the Bayes' rule is $\Pr(p)\Pr(+|p)$, where $\Pr(p)$ is the prevalence (4%), and $\Pr(+|p) = 1 - \Pr(-|p) = 1 - 0.03 = 0.97$. That is, the numerator is 0.0388. We can easily verify that the denominator is 0.0388 + 0.048. Multiplying the population number with both the numerator and denominator, we see that the Bayes' rule simply tallies the number of true positives and total positives.
- When the test is imperfect, we are uncertain about the result. The *posterior probability* $\Pr(p|+)$, the probability of present after observing a positive result, is used to summarize the uncertainty we have on the result. The uncertainty, however, is a function of the prior and the two probabilities characterizing the performance of the test. In the numerical example here, if the prior is 1%, the posterior would be $\Pr(p|+) = 0.164$. This outcome is expected as the number of true positives is now a smaller fraction of the total number of positives (Figure 1(iv)-(vi)).

B. The meaning of a prior

- Different people may have different prior probabilities for the same event. This is likely because they use different references. For example, when a doctor tests for a disease in a public health exhibition at a county fair, her prior should be the prevalence of the disease in the population because she would consider the test subject a random sample from the population. To the patient, the general population may not be a good reference because he knows more about himself. In this case, he knows which risk factors apply to him. As a result, the relevant population would be people with the same risk factors (e.g., smokers).
- The meaning of the prior to the patient and to the doctor may be the same (e.g., prevalence of the disease in a population). But deciding which population to use to form the prior requires more information. The subjective nature of a Bayesian prior is often criticized. We contend that a prior is simply a means for scientists to properly sort out the relevant facts/information using their knowledge of their study system. In this regard, when we define the prior probability as a degree of belief, we are really trying to make use of all our knowledge to ensure the outcome is most relevant.
- When we cannot confidently identify a sub-population for inference, we must step back to a larger population. The resulting prior is likely less relevant. As a result, the estimated posterior probability is less accurate. For example, in the numerical example in SI-A., the posterior probability of present given a positive results is less than 0.5. But the test result puts the snake into a smaller population (the 388+480 would-be positive snakes). If we choose to conduct a follow-up test, our prior would be the posterior from the first test. This iterative process is an appealing characteristic of the Bayesian method for many applied scientists.

C. Inverse Bayes Formulae (IBF)

IBF can be derived directly using the Bayes' rule or, more parsimoniously, using the joint distribution formula: $\Pr(X,Y) = \Pr(Y|X)\Pr(X) = \Pr(X|Y)\Pr(Y)$, which yields

$$\Pr(Y) = \frac{\Pr(Y|X)}{\Pr(X|Y)}\Pr(X)$$

We set X to be the variable of the state of the world (taking values p or a) and Y the variable of test results ($+$ or $-$). Because $\Pr(Y=+)+\Pr(Y=-) = 1$, $\Pr(Y=-) = \frac{\Pr(Y=-|X)}{\Pr(X|Y=-)}\Pr(X)$, and $\Pr(Y=+) = \frac{\Pr(Y=+|X)}{\Pr(X|Y=+)}\Pr(X)$, we have

$$\frac{\Pr(Y=+|X)}{\Pr(X|Y=+)}\Pr(X) + \frac{\Pr(Y=-|X)}{\Pr(X|Y=-)}\Pr(X) = 1 \qquad [4]$$

From equation Eq. (4), we can express the quantity of interest $\Pr(X=p)$ as:

$$\Pr(X=p) = \frac{1}{\frac{\Pr(Y=+|X=p)}{\Pr(X=p|Y=+)} + \frac{\Pr(Y=-|X=p)}{\Pr(X=p|Y=-)}}$$

or, using the notation in SI-A.:

$$\Pr(p) = \frac{1}{\frac{\Pr(+|p)}{\Pr(p|+)} + \frac{\Pr(-|p)}{\Pr(p|-)}} \qquad [5]$$

Assuming that the false negative probability ($\Pr(-|p)$) is a known quantity characterizing the test, we then know two of the four probabilities ($\Pr(-|p)$ and $\Pr(+|p) = 1 - \Pr(-|p)$). The other two probabilities ($\Pr(p|-), \Pr(p|+)$) can be estimated if the true state of the world of an individual test subject can be ascertained. Suppose that we test N individuals and record the test result. In the follow up test we determine the true status of these individuals and record the outcome in a two by two table:

	$X=p$	$X=a$	total
$y=+$	n_{11}	n_{12}	N_1
$y=-$	n_{21}	n_{22}	N_2
total	M_1	M_2	N

From the above table, we have $\Pr(p|-) = n_{21}/N_2$ and $\Pr(p|+) = n_{11}/N_1$. In the rattlesnake example, we can ascertain the status of the tested snakes by keeping them in the lab for a period of time to observe which ones eventually succumb to the fatal disease.

D. Bayesian inference

- Using Bayesian statistics to estimate an unknown variable consists of three steps.
  (a) Propose a statistical model that describes the data. This model includes the unknown variable as a parameter. For example, the statistical model we use to estimate the prevalence of SFD in the Michigan rattlesnake population is a binomial model.
  (b) Using the statistical model, we derive the likelihood function of the data – the likelihood of observing the data if the proposed model is correct. If the test is perfect, that is, both false positive and false negative rates are 0, the prevalence is the probability of observing a positive result. The statistical model of observing y positive in n snakes is

$$\Pr(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \quad [6]$$

This is the likelihood of observing the data if the prevalence is θ. In classical statistics, the estimated parameter $\hat{\theta}$ is the value that maximizes the likelihood.

(c) Estimating the posterior distribution of θ using the Bayes' rule of a continuous variable:

$$\pi(\theta|y) = \frac{\pi(\theta)\Pr(y|\theta)}{\int_\theta \pi(\theta)\Pr(y|\theta)d\theta} \quad [7]$$

where $\pi()$ represent a probability density function. $\pi(\theta)$ is the prior distribution, representing the uncertainty we have about θ and $\pi(\theta|y)$ is the posterior distribution of θ after observing data.

- The derivation of the posterior distribution is often the difficult part of Bayesian inference because of the integral in the denominator. In some cases, the derivation can be simplified if we choose a suitable prior distribution. For example, when we choose the beta distribution as the prior for θ, the prior density (with parameters α and β) is $\pi(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$, the posterior is

$$\begin{aligned}\pi(\theta|y) &= \frac{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_\theta \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{(n-y)+\beta-1}\end{aligned} \quad [8]$$

The posterior is the beta distribution with parameters $y+\alpha$ and $(n-y)+\beta$. The mean of the posterior is $\frac{y+\alpha}{n+\alpha+\beta}$. From this distribution, we can also calculate the credible interval to summarize the uncertainty.

- When the test is imperfect and we know the rate of false positive (f_p) and rate of false negative (f_n), the model becomes more complicated because the probability of observing a positive result is now $\theta^\star = \theta(1-f_n)+(1-\theta)f_p$. The statistical model of the data is still the binomial distribution, but the probability of observing a positive result is now $\theta^\star$. As a result, the posterior distribution is:

$$\begin{aligned}\pi(\theta|y) \quad \propto \quad & \theta^{\alpha-1}(1-\theta)^{\beta-1}\times \\ & (\theta(1-f_n)+(1-\theta)f_p)^y \times \\ & (1-\theta(1-f_n)-(1-\theta)f_p)^{n-y}\end{aligned} \quad [9]$$

E. Bayesian computation

- In SI-D. we derived two posterior distributions of the parameter θ. One is summarized by a standard probability distribution, of which we know how to derive needed statistics to summarize the uncertainty about the parameter. The other is an algebraic expression that cannot be represented by a known form of probability distribution. Because θ is the only unknown parameter, we can graphically draw the posterior distribution on a two dimensional space. By numerically re-scaling the curve such that the area under the curve is 1, we have a graphical representation of the probability distribution (Figure 3), from which we can draw inference about the parameter. When one of the two error rates (e.g., the false positive rate f_p) is unknown, the problem is to estimate the joint distribution of the two unknowns. We can still use the same numerical approximation method and graphically present the posterior using a contour plot (Figure 4). The same computational framework can be used if both error rates are unknown. But graphical display of the distribution is no longer feasible. Frequently, we use the modern Bayesian computational method for this type of problem. Computational details of this paper can be found at the lead author's GitHub page.