

Bayes and diagnostic testing

Emmanuel Lesaffre^a, Niko Speybroeck^b, Dirk Berkvens^{b,*}

^a Biostatistical Centre, Catholic University of Leuven, Leuven, Belgium

^b Institute of Tropical Medicine, Department of Animal Health, Nationalestraat 155, B-2000 Antwerp, Belgium

Abstract

Interpretation of the result of a diagnostic test depends not only on the actual test result(s) but also on information external to this result, namely the test's sensitivity and specificity. This external information (also called prior information) must be combined with the data to yield the so-called updated, posterior estimates of the true prevalence and the test characteristics. The Bayesian approach offers a natural, intuitive framework in which to carry out this estimation process. The influence of the prior information on the final result may not be ignored. Guidance for the choice of prior information not in conflict with the data can be obtained from a set of statistics and indices (DIC, p_D , Bayes-p).

© 2007 Elsevier B.V. All rights reserved.

Keywords: Diagnostic test; True prevalence; Sensitivity; Specificity; Bayesian

1. Introduction

The Reverend Thomas Bayes lived from 1702 to 1761. He was the son of a Nonconformist minister and was educated privately, rumour having it that one of his teachers was the French statistician Abraham de Moivre, who had previously 'discovered' the normal distribution. Thomas Bayes in turn was also ordained minister. He remained interested in mathematics, probability theory and statistics throughout his life. We now know Bayes' ideas on statistics mainly through two posthumously published papers (Bayes, 1763; Price, 1764). The first paper contains the solution to a problem posed by de Moivre in his *Doctrine of Chances* in 1718, hence the paper's title.

Bayes is thought to be the first to use probability theory inductively. He developed the mathematical basis for probability inference, which is a method to

calculate the probability that an event will occur in the future, based on the frequency with which this event has occurred in the past. According to Bayes, all quantities involved in inference belong to two kinds: those known and those unknown to the person making the inference. The first kind will enter the inference with its known (accepted) values. The second kind enter the equation as probability distributions reflecting expert opinion, making the Bayesian view thus a completely new way of looking at statistics (Johnson and Kotz, 1997).

Bayes' views went largely unnoticed being cited without challenge (by, among others, Laplace) until Boole questioned them (Boole, 1854). The controversy continued until they were completely discredited by early 20th century statisticians. It was only during the 1950s that Bayesian statistics were rediscovered through the efforts of statisticians, econometricians and physicians, increased computational power undoubtedly aiding this resurrection. The statisticians (among them De Finetti, Jeffreys, Savage and Lindley) developed a complete method of statistical inference based on Bayes' theorem (Bolstad, 2004).

* Corresponding author. Tel.: +32 3 2476393; fax: +32 3 2476268.
E-mail address: dberkvens@itg.be (D. Berkvens).

2. Bayes' theorem

Bayes' theorem or Bayes' formula introduces inverse probability to be used when calculating the probability of antecedent events based on the occurrence of consequent events. Think of the antecedent event as the point in time when the animal became infected and the consequent event as the result of a diagnostic test carried out at a later point in time. 'Infected' is used in the most general sense of the word and may be replaced by 'infested', 'diseased', 'carrier' or whatever term to distinguish this animal (a case) from a non-case.

The actual Bayes' formula is:

$$P(\text{Ante}|\text{Cons}) = \frac{P(\text{Cons}|\text{Ante}) \cdot P(\text{Ante})}{P(\text{Cons})}$$

with Ante = having become infected; Cons = positive test result or, in terms of infection and test result:

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

where D^+ refers to 'being infected' and T^+ to 'a positive test result'. The formula thus calculates the probability that a particular animal is infected given a positive test result. In the latter form, it becomes clear that this formula is what is commonly known in epidemiology as the predictive value of a positive test result and it is indeed exactly what inverse probability signifies: what is the probability that an animal was infected previously, given a positive test result now? Bayes' theorem is thus in fact a commonly used tool when having to decide whether or not to classify an animal as infected or not.

In order to appreciate how Bayes' formula calculates this after-test probability, we can replace the various probabilities by better known terms.

As mentioned before $P(D^+|T^+)$ refers to the predictive value of a positive test result (PV^+). $P(T^+|D^+)$ represents the probability of a positive test result given an infected animal, a conditional probability known as the sensitivity of the diagnostic test (Se). $P(D^+)$ is the probability that any randomly chosen animal from a population is infected, a population parameter known as the true prevalence of infection (p). Lastly, $P(T^+)$ stands for the probability of obtaining a positive test result, irrespective of the actual status of the animal. This probability is the proportion of positive test results obtained when testing a certain number of animals and is known as the apparent prevalence, or seroprevalence or laboratory prevalence (p'). The probability of a positive test result (p') is the sum of

the probability of a positive test result given an infected animal (prevalence times sensitivity) and the probability of a positive test result given an uninfected animal (probability of not being infected, or $1 - \text{prevalence}$, times lack of specificity, or $1 - \text{Sp}$). PV^+ thus becomes:

$$PV^+ = \frac{\text{Se} \cdot p}{p'}$$

$$PV^+ = \frac{\text{Se} \cdot p}{\text{Se} \cdot p + (1 - \text{Sp}) \cdot (1 - p)}$$

3. Prior information, data likelihood and posterior information

Prior information, data likelihood and posterior information are terms commonly used in Bayesian analysis (Congdon, 2003) and they are best explained using the previous example of testing an infected animal.

Prior information refers to our knowledge about the parameter of interest (i.e. the probability that an animal is infected) without the test being performed. Our best guess concerning the probability of infection when selecting an animal at random from a population is the prevalence and this parameter represents our prior knowledge (or belief, or degree of belief) about the probability of infection in any animal in that population. This prior information can be an exact value, a range of values, an expert opinion, a probability distribution,...

The data likelihood is embodied by the (conditional) probability of a positive test result, given in the first place the test sensitivity, but also the prevalence of the infection and the lack of specificity of the test.

Combining the prior information and the data likelihood yields the posterior information, namely the updated probability of the tested animal being infected, given the fact that it yielded a positive test result. This new, updated probability corresponds to our update or post-test belief in the chances that the animal is truly infected.

4. Information inherent to the data and information external to the data

The previous section already hinted that some information, essential for arriving at a meaningful decision when applying a diagnostic test to an animal, is actually not contained in the test results (i.e. is external to the data).

This is easier to understand if we extract the prevalence (p) from the formula for apparent prevalence

$$p' = p \cdot \text{Se} + (1 - p) \cdot (1 - \text{Sp})$$

thereby obtaining the Rogan–Gladden estimator (Rogan and Gladden, 1978) of the true prevalence

$$p = \frac{p' + \text{Sp} - 1}{\text{Se} + \text{Sp} - 1}$$

We need to estimate three parameters if we want to estimate the true prevalence, namely the apparent prevalence (p'), the test sensitivity (Se) and the test specificity (Sp). The data (the test results) contain only information about the first parameter. The apparent prevalence is estimated as the number of positive test results divided by the number of animals tested. Information about the two other parameters is not present in the data and has to be supplied independently from and external to the data. The value of the test sensitivity and test specificity has to be obtained either from other data (experimental results, other studies, . . .) or, in the absence of data, from expert opinion.

A Bayesian analysis allows us to combine this external information (i.e. prior information) with the data to yield an estimate. It immediately becomes clear that our conclusion (prevalence estimate or predictive value of a test) is the result of both the data and the prior information, in other words the result of our prior beliefs and the new evidence. Bayesian analysis allows us to assign probability distributions to our prior beliefs and combine these with the data likelihood to yield a posterior probability distribution representing our updated belief.

The important point here is that test sensitivity and test specificity actually also have become variables. In the classic approach test sensitivity and specificity are fixed parameters and the true prevalence is calculated from them, using the above mentioned Rogan–Gladden formula. The only variable in the system is the apparent prevalence. But, test sensitivity and specificity are in fact variables, their value being independent of the data at hand and the value used in the equation being based entirely on expert opinion: the person (or persons) transforming the apparent prevalence into a true prevalence decide what values to use, it is their chosen values that determine the outcome. The fact that test sensitivity and specificity are variables is completely in line with the Bayesian spirit, where indeed a distinction between variables and parameters ceases to exist and where instead we deal with known and unknown variables, information about the former being found in

the data, the latter's value being decided upon by the experts. The fact that this is in truth also the case in real life is often not realised, overlooked or even conveniently ignored.

The bottom-line is that any change in the value of either test sensitivity or test specificity inserted into the Rogan–Gladden equation immediately alters the estimate of the true prevalence, i.e. the true prevalence estimate is conditional not only on the data, but also on the expert opinion.

Another way of looking at this problem is to observe that the number of parameters to be estimated from the data exceeds the degrees of freedom in the data. For instance in the one-test case, we have to estimate three variables (the true prevalence, the test sensitivity and the test specificity), but we can only estimate a single variable (we can compute a single unknown as we have only one equation). This is called an underspecified system. Unfortunately, it can be shown that the estimation of true prevalence is always based on an underspecified system of equations, meaning that we invariably have to estimate more variables than the number of equations at hand.

5. Multi-testing

A possible solution to the problem in the previous section lies in so-called multi-testing, i.e. every animal is subjected to different diagnostic tests.

Multi-testing as such is no direct solution to the under-specification problem. Applying more than one test to the same animal in fact introduces extra variables, required to cater for the interdependence of the different tests, i.e. an infected animal yielding a positive test result for one serological test is more likely than not to test positive in another serological test and this conditional interdependence must be included in the model (Branscum et al., 2005). Several approaches have been advanced to try and circumvent this problem, the most widely used being the one proposed by Hui and Walter (1980). Several authors (Gustafson, 2005; Johnson et al., 2000; Toft et al., 2005) have voiced their concerns about this approach, pointing out that several assumptions made (e.g. constancy of test sensitivity and specificity, independence of tests) are very strong and not necessarily correct.

Multi-testing, using Bayesian or frequentist statistics, has received ample attention recently (Branscum et al., 2005; Enoe et al., 2000; Johnson et al., 2000). However, as pointed out by Gustafson (2005), there is no universal ready-made solution to overcome the problem of under-specification. There is no escaping

from the fact that external information is required to estimate the true prevalence and it must of course be understood that the external information codetermines the posterior estimates.

The choice of prior (i.e. external) information thus assumes a central role when interpreting diagnostic test results and every effort should be made to verify that the prior probability distributions assigned to the different variables are not in conflict with the data (Berkvens et al., 2006). The method proposed in this paper (Berkvens et al., 2006) uses a set of statistics and indices (DIC, p_D , Bayes-p) to allow the choice of prior information in accordance with the data. This method alerts the user if the prior information is not in accordance with the data, permitting the choice of prior information that is not in contradiction with the data. It is obvious that the ultimate estimates (e.g. true prevalence) remain the result of the data and the prior information. This approach has been utilised in several instances, among them the estimation of the true prevalence of porcine cysticercosis and calf giardiasis (Dorny et al., 2004; Geurden et al., 2004).

6. Discussion

Realising the need of external information is a first step towards correct interpretation of diagnostic test results. The impossibility to estimate the true prevalence in the absence of a gold-standard test (sensitivity and specificity both equal to unity) without the addition of information not present in the data must be appreciated by everybody relying on the correct interpretation of test results. This external information is always at least partly expert opinion: even when solid data exist with regards to (e.g.) test sensitivity and test specificity, somebody must still decide that this information is pertinent for the test results to be interpreted. The influence of this prior external information on the posterior estimates must also be remembered. The Bayesian approach provides us with a natural, intuitively correct framework that can be used to guide us through the process of test result interpretation. The judicious use of the appropriate statistics and indices allow the user to sieve out combinations of expert opinions that are in conflict with the data. However, it must always be understood that

absence of conflict between prior information and data is no guarantee that the best (or worse the only) combination of priors has been selected and applied. The final estimate remains the result of the data and the external information.

References

- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc.* 53, 370–418.
- Berkvens, D., Speybroeck, N., Praet, N., Adel, A., Lesaffre, E., 2006. Estimating disease prevalence in a bayesian framework using probabilistic constraints. *Epidemiology* 17, 145–153.
- Bolstad, W., 2004. *Introduction to Bayesian Statistics*. John Wiley & Sons, Inc, Hoboken.
- Boole, G., 1854. *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*. Walton and Maberley, London.
- Branscum, A., Gardner, I., Johnson, W., 2005. Estimation of diagnostic-test sensitivity and specificity through bayesian modelling. *Prev. Vet. Med.* 68, 145–163.
- Congdon, P., 2003. *Applied Bayesian Modelling*. John Wiley & Sons, Ltd, Chichester.
- Dorny, P., Phiri, I., Vercruyse, J., Gabriel, S., Willingham, A., Brandt, J., Victor, B., Speybroeck, N., Berkvens III, D., 2004. A bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. *Int. J. Parasitol.* 34, 569–576.
- Enoe, C., Georgiadis, M.P., Wo, J., 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true diseases state is unknown. *Prev. Vet. Med.* 45, 61–81.
- Geurden, T., Claerebout, E., Vercruyse, J., Berkvens, D., 2004. Estimation of diagnostic test characteristics and prevalence of giardia duodenalis in dairy calves in Belgium using a bayesian approach. *Int. J. Parasitol.* 34, 1121–1127.
- Gustafson, P., 2005. The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. *Stat. Med.* 24, 1203–1217.
- Hui, S., Walter, S., 1980. Estimating the error rates of diagnostic tests. *Biometrics* 36, 167–171.
- Johnson, N., Kotz, S., 1997. *Encyclopedia of Statistical Sciences*, vol. 1. John Wiley & Sons, Inc, New York.
- Johnson, W., Gastwirth, J., Pearson, L., 2000. Screening without a “gold standard”: the hui-walter paradigm revisited. *Am. J. Epidemiol.* 153, 921–924.
- Price, R., 1764. A demonstration of the second rule in the essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc.* 54, 296–325.
- Rogan, W., Gladen, B., 1978. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* 107, 71–76.
- Toft, N., Jorgensen, E., Hojsgaard, S., 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Prev. Vet. Med.* 68, 19–33.