The Veterinary Journal 202 (2014) 204–207

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

Personal View

Bayesian versus frequentist methods for estimating true prevalence of disease and diagnostic test performance

Moussa Sanogo [a,b,c], Emmanuel Abatih [b], Claude Saegerman [c,*]

[a] LANADA/Central Veterinary Laboratory of Bingerville, P.O. Box 206, Bingerville, Ivory Coast
[b] Unit of Epidemiology and Biostatistics, Department of Biomedical Sciences, Institute of Tropical Medicine, Nationalestraat, 155, B-2000 Antwerp, Belgium
[c] Research Unit of Epidemiology and Risk Analysis Applied to Veterinary Sciences (UREAR-ULg), Fundamental and Applied Research for Animals and Health, Faculty of Veterinary Medicine, University of Liège, Boulevard de Colonster 20, B42, B-4000 Liège, Belgium

ARTICLE INFO

Article history:
Accepted 4 August 2014

Keywords:
Bayesian methods
Frequentist methods
Diagnostic test performance
True prevalence
Sensitivity
Specificity

Introduction

As the two main schools of statistical reasoning through which inference to the population is made by analysing data and incorporating uncertainty of measures, Bayesian and frequentist philosophies have been used to estimate diagnostic test performance and the true prevalence of diseases. However, controversies exist between the two philosophies, such as the use of fixed parameter values in the frequentist approach or the inclusion of prior information in the Bayesian approach. So, is the philosophical debate between these two approaches still relevant for such practical questions?

The Bayesian philosophy arose from a statement made by the Reverend Thomas Bayes (1702–1761), a British mathematician and theologian, who was the first to apply statistical probability inductively. According to Bayes, '*all forms of inference are based on the validity of their premises*' and '*no inference can be known with certainty*' (Thrusfield, 2005). In 1814, the French mathematician, Simon-Pierre Laplace published a mathematical description based on Bayes's idea (Gelman et al., 2004). In the Bayesian philosophy, scientific observations do not exist in a vacuum and information available prior to making a series of observations influences the interpretation of those observations (Thrusfield, 2005).

Bayesian analysis can be regarded as a process of adjusting and updating the likelihood of an event based on data. Thus, population parameters, such as sensitivity (Se) and specificity (Sp), are assumed to have a probability distribution representing our prior knowledge of their values. This information is combined with observed factual field data in a model for estimation (Speybroeck et al., 2012a). For Bayesians, a parameter is assumed to have an intrinsic probability distribution with a 95% credibility interval (Gardner, 2002). Thus, Bayesian principles are often applied in order to estimate disease prevalence and test characteristics, especially when there is no gold standard (Enøe et al., 2000; Branscum et al., 2005; Rutjes et al., 2007; Meyer et al., 2009).

The frequentist philosophy emerged in the 20th century with the works of Fisher (1922) and Neyman and Pearson (1928a,b), who enunciated the concept of relative frequency (Vallverdú, 2008). This concept sustains the idea that a probability is a frequency determined from an experiment repeated a large number of times. Frequentist statisticians attempt to draw conclusions by focussing primarily on results obtained from experiments or samples. In the frequentist reasoning, a parameter is a fixed value with a 95% confidence interval derived from the sample. It is assumed that this 95% confidence interval would contain the true value of the parameter 95% of the time if the estimation were repeated a large number of times.

Therefore, Bayesian philosophical methods are based on the idea that unknown quantities, such as population means or proportions, have a probability distribution that expresses our prior knowledge or belief about such quantities, *before* we add the knowledge gained from observational data. Bayesian inference considers the data to be fixed and parameters to be random, because they are unknown. In frequentist

Please note that the content in this Personal View article has not been subject to peer-review. The views expressed in this Personal View are entirely those of the author(s) and do not necessarily reflect those of the editorial team, or Elsevier.

* Corresponding author. Tel.: +32 4 3664579.
E-mail address: claude.saegerman@ulg.ac.be (C. Saegerman).

http://dx.doi.org/10.1016/j.tvjl.2014.08.002
1090-0233/

methods, prior knowledge is apprehended differently and population means or proportions are considered as fixed values (Bland and Altman, 2002). Frequentist inference considers the unknown parameters to be fixed and the data to be random.

Bayesian and frequentist methods have been published to handle a variety of situations in which diagnostic tests are evaluated. In this article, we consider the requirements, limitations and controversial points of the methods proposed for estimating test performance and the true prevalence of disease, using the case where one test or a combination of two imperfect diagnostic tests is applied in the absence of an appropriate gold standard.

Estimating the true prevalence of disease and diagnostic test performance with imperfect tests

The ability of a diagnostic test to distinguish correctly truly diseased from non-diseased individuals, when applied to a randomly chosen population, is necessary so as to understand the epidemiology of the disease, to implement disease control programmes and to evaluate new diagnostic tests (Greiner and Gardner, 2000; Lewis and Torgerson, 2012). Mathematically, the estimation of test performance parameters is essentially the same question as estimating true prevalence (Lewis and Torgerson, 2012). The true prevalence (the proportion of truly diseased individuals in the population of interest) is also an essential parameter required to appraise the impact of a disease in a population of interest and to prevent biased estimations of disease burden (Dohoo et al., 2003; Speybroeck et al., 2012a).

The accuracy of estimation of true prevalence depends on the performance parameters of the test(s) to be applied (Ihorst et al., 2007). Among performance indicators of a diagnostic test, Se and Sp are the most commonly used. Test Se or Sp indicates the probability that a truly infected or uninfected individual yields a positive or negative test result, respectively. Ideally, Se and Sp values for a given test should be estimated from a reference population with a clearly identified status determined by historical (accurate) information or, more commonly, by a relevant gold standard (Se = 1 and Sp = 1) that is able to discriminate infected/diseased individuals from uninfected/non-diseased individuals in a population (Dohoo et al., 2003). When such a perfect test exists, an estimation of performance parameters of the new test, as well as true prevalence, can be done easily (Rogan and Gladen, 1978).

In practice, such a test is hardly ever available, given that the diagnostic performance of a test is influenced by a number of endogenous and exogenous factors (Rutjes et al., 2007). As an alternative, a combination of multiple imperfect tests (Se < 1 and/or Sp < 1) may be used to estimate disease parameters (Black and Craig, 2002). With multiple tests, overall misclassification errors are reduced and are expected to be lower than with a single imperfect test.

For example, the isolation and identification of *Brucella* spp. is considered as the reference standard method; a positive test result provides an unequivocal diagnosis of a positive case of brucellosis (World Animal Health Organisation, 2009). However, these methods are not always feasible in diagnostic investigations. Therefore, diagnosis must be based on imperfect serological methods, such as the Rose Bengal test (RBT) and the indirect ELISA (iELISA), which are the two OIE prescribed tests for trade and are commonly used in combination for the diagnosis of brucellosis (Nielsen, 2002; Saegerman et al., 2004; World Animal Health Organisation, 2009; Godfroid et al., 2010; Sanogo et al., 2013).

Estimation of true disease prevalence and test characteristics with combined imperfect tests poses challenges, including (1) potential misclassification errors, (2) possible dependence between tests, and (3) sparseness of data (Cowling et al., 1999; Dohoo et al., 2003; Messam et al., 2008). Both Bayesian and frequentist approaches have been proposed to tackle these challenges.

Estimation with a single test

In the simple case, where a single imperfect diagnostic test is applied in a population of interest, a total of three parameters must be estimated, whatever the method, namely, Se, Sp and true prevalence. In this case, the apparent prevalence (the proportion of positive test results) is the only information given by the data. From a frequentist perspective, estimation can be done only if fixed external information is provided on the values of Se and Sp, but this is difficult, since test properties are known to be context-specific and cannot realistically be assumed to be fixed and known in advance, such as the values given by the manufacturer of a test (Thrusfield, 2005).

As far as external information has to be included for estimations, Bayesian methods seem to be more helpful in obtaining acceptable and realistic results, since they offer the possibility of including the known uncertainty with respect to diagnostic test characteristics, while testing whether data conflict with prior information (Joseph et al., 1995; Berkvens et al., 2006; Speybroeck et al., 2012b). However, the accuracy of Bayesian estimates is dependent on the availability and quality of prior knowledge, which may be a limiting factor and may also conflict with frequentist philosophy.

Estimation with more than one test

When a combination of at least two tests is used, the test results for a given individual could be interpreted either in series (only animals that test positive to both tests are considered to be test positive) or in parallel (animals that test positive to one test, to the other test or to both tests are considered to be test positive) (Black and Craig, 2002). A combination of tests may also result in dependence or correlation between the test results. As a consequence, either conditional independence or conditional dependence assumptions need to be made for accurate estimation of disease prevalence and test properties (Jones et al., 2010).

Conditional independence implies that the results of the second test (T2) do not depend on whether the results of the first test (T1) are positive or negative among infected (or uninfected) individuals (Enøe et al., 2000; Gardner et al., 2000). If we consider the skin test or the iELISA, two tests referred to above for the diagnosis of brucellosis, conditional independence is likely to exist in relation to their respective targets (cellular response for the skin test and humoral response for iELISA), especially in a low prevalence context (Saegerman et al., 1999). In this case, calculation of test Se and Sp will depend mainly on the testing strategy adopted (in parallel or in series) (Dohoo et al., 2003).

Mathematically, assumptions such as conditional independence and a constant prevalence over sub-populations are needed to estimate prevalence (Enøe et al., 2000). These assumptions are necessary so as to reduce the number of unknown parameters to be estimated (Berkvens et al., 2006). Gart and Buck (1966) and Staquet et al. (1981) proposed frequentist methods assuming conditional independence between a new test and a reference test with known Se and/or Sp. However, test Se (stage of infection) and Sp (similar immunogenic component) values are influenced by the characteristics of the population in which the test is applied (Saegerman et al., 2004; Berkvens et al., 2006) and cannot be considered as intrinsic constant and known parameters (Thrusfield, 2005). Moreover, assuming fixed values might not be realistic, since many factors, such as the presence of cross-reacting agents (Saegerman et al., 2004) and low infection pressure, may influence test parameter values (Speybroeck et al., 2012b).

Hui and Walter (1980) proposed another major frequentist method to deal with the case where Se and Sp values of the reference test are unknown. In addition to an assumption of conditional independence, this approach required testing at least two populations with distinct prevalences of disease, but constant Se and Sp (Hui and Zhou, 1998; Enøe et al., 2000; Dohoo et al., 2003). The approach was extended to