Personal view

Bayesian versus frequentist methods for estimating disease true prevalence 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 (ITM), 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 & Health, Faculty of Veterinary Medicine, University of Liege (ULg), Boulevard de Colonster 20, B42, B-4000 Liege, Belgium*

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

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 for estimation of diagnostic test performance and the true prevalence of diseases. However, some controversies exist in this estimation between these two philosophies like (e.g.) the use of fixed parameters values in frequentist approach or the inclusion of prior information in Bayesian approach. 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 that '*no inference can be known with certainty*' (Thrusfield, 2005). In 1814, the French mathematician, Simon-Pierre Laplace published a mathematical description based on idea of Bayes (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 often are applied to estimate disease prevalence and test

47 characteristics, especially when there is no gold standard, in veterinary or human health (Enøe
48 et al., 2000; Branscum et al., 2005; Rutjes et al., 2007; Meyer et al., 2009).
49
50 The frequentist philosophy emerged in the 20th century with the works of Fisher
51 (1922) and Neyman and Pearson (1928), who enunciated the concept of relative frequency
52 (Vallverdu, 2008). This concept sustains the idea that a probability is a frequency determined
53 from an experiment repeated a large number of times. Frequentist statisticians attempt to draw
54 conclusions by focussing primarily on results obtained from experiments or samples. In the
55 frequentist reasoning, a parameter is a fixed value with a 95% confidence interval derived
56 from the sample. It is assumed that this 95% confidence interval would contain the true value
57 of the parameter 95% of the time if estimation were repeated a large number of times.
58
59 Therefore, Bayesian philosophical methods are based on the idea that unknown
60 quantities, such as population means or proportions, have a probability distribution that
61 expresses our prior knowledge or belief about such quantities, before we add the knowledge
62 gained from observational data. Bayesian inference considers the data to be fixed and
63 parameters to be random, because they are unknown. In frequentist methods, prior knowledge
64 is apprehended differently and population means or proportions are considered as fixed values
65 (Bland and Atlman, 2002). Frequentist inference considers the unknown parameters to be
66 fixed and the data to be random.
67
68 Both Bayesian and frequentist methods have been published to handle a variety of
69 situations in which diagnostic tests are evaluated. In this personal view, we comment on
70 requirements, limitations and controversial points of proposed methods for estimating test
71 performance and the true prevalence of disease, through the case where one test or a

combination of two imperfect diagnostic tests is used 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 correctly distinguish truly diseased from non-diseased individuals when applied to a randomly chosen population is required to understanding 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, 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 to appraise the impact of a disease in a population of interest and to prevent biased estimation 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 non-infected) individual yields a positive (or a negative) test result. 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), which is able to discriminate infected/diseased individuals from non-infected/non-diseased individuals in a population (Dohoo et al., 2003). When such a perfect test exists, 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 for estimation of disease parameters (Black and Graig, 2002). With multiple tests, the overall misclassification errors are reduced and are expected to be lower than with a single imperfect test.

As an example, isolation and identification of *Brucella* spp is considered as the reference standard method, and a positive test result provides an unequivocal diagnosis of a positive brucellosis case (OIE, 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 proscribed tests for trade and are commonly used in combination for the diagnosis of brucellosis (Nielsen, 2000; Saegerman et al., 2004; OIE, 2009; Godfroid et al., 2010; Sanogo et al., 2013).

Estimation of true disease prevalence and test characteristics with combined imperfect tests implies to deal with various 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 have to be estimated, whatever the method: 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 be realistically assumed to be fixed and known in advance (Thrusfield, 1995), for example, the given values by the manufacturer of a test.

As far as external information has to be included for estimation, Bayesian methods seem to be more helpful in obtaining acceptable and realistic results, since they offer the possibility to include the known uncertainty on 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 the quality of prior knowledge, which may be a limiting factor and also constitute a source of controversy with frequentist philosophy.

Estimation with a more than a single 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 testing positive to both tests are considered to be test positive) or in parallel (animals that tested positive to one test, to the other test or to both tests are considered to be test positive) (Black and Graig, 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 non-infected) individuals (Gardner et al., 2000; Enøe et al, 2000). If we consider the skin test or the iELISA, two tests for the diagnosis of brucellosis, conditional independence is likely to exist considering 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 (in parallel or in series) adopted (Dohoo et al., 2003).

Mathematically, assumptions such as conditional independence and a constant prevalence over sub-populations, are needed to estimate the prevalence (Enøe et al., 2000). These assumptions are necessary 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 known to be under the influence of 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, 1995). 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).

Another major frequentist method was proposed by Hui and Walter (1980) to deal with the case where Se and Sp values of the reference test are unknown. In addition to

conditional independence assumption, this latter approach required testing results from 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). This approach was extended to cover other settings, including cases with more than two tests and multiple populations (Walter and Irwig, 1988; Johnson et al., 2001). Accuracy of estimates with these methods also relies on the assumption of a large sample size (Enøe et al., 2000; Pouillot et al., 2002; Berkvens et al., 2003; Pouillot, 2003). Toft et al. (2005) provide a useful overview of possible pitfalls when using this paradigm, especially the conditional independence assumption, which is not always satisfied in practice (Gardner et al., 2000; Dendukuri and Joseph, 2001; Branscum et al., 2005; Berkvens et al, 2006).

Testing situations handled by the frequentist models of Gart and Buck (1966) and the case of unknown Se and Sp already covered by the model of Hui and Walter (1980) were also examined under the Bayesian framework. Joseph et al. (1995) proposed a Bayesian model for estimation with no constrained parameters and assuming conditional independence. Numerically, this model appeared to be approximately equivalent to the frequentist approach (Dendukuri and Joseph, 2001). Nevertheless, even if estimation was possible with this latter model, inclusion of information on the uncertainty of parameters to be determined is required to get realistic and meaningful estimates (Enøe et al., 2000).

Conditional dependence particularly occurs when combined tests target a similar biological phenomenon, such as presence of immunoglobulins (Igs) (Gardner et al., 2000; Dendukuri and Joseph, 2001). Thus, conditional dependence is likely to exist between the RBT and iELISA, two assays targeting the similar anti-*Brucella* antibodies. In fact, RBT detects the presence of IgG_1 (IgG_2 and IgM also have some agglutination activity), while the

196 iELISA targets IgG_1 and/or IgG_2, depending on the conjugate used (Nielsen, 2002;
197 Saegerman et al., 2004 and 2010; Sanogo et al., 2013). In this scenario, calculation of test Se
198 and Sp under conditional independence is adjusted with the inclusion of the covariance factor
199 expressing the extent of the dependence among positive and negative results, and by taking
200 the testing strategy into account (Dohoo et al., 2003).
201
202 When dependence is present, adjustment of estimations should be done by considering
203 biological and technical mechanisms giving rise to the test results and by including the extent
204 of the dependence between them (Pepe and Janes, 2007). With two correlated tests, a total of
205 seven parameters have to be estimated instead of five under conditional independence (e.g.
206 two sensitivities, two specificities, two covariances and the true prevalence) and the
207 dependence needs to be accounted for (Berkvens et al., 2006; Praet et al., 2006). Some
208 frequentist methods require the application of at least two tests to allow estimation of
209 parameters of interest (Dendukuri and Joseph, 2001). Such an approach might be impractical,
210 when tests are expensive, time consuming or invasive.
211
212 Instead of using results from at least two tests to allow estimation of disease
213 parameters, Bayesian modelling offers an alternative option to get the estimates of the true
214 prevalence of disease and test Se and Sp, while accounting for conditional dependence (Qu
215 and Hagdu, 1998; Gardner et al., 2000; Dendukuri and Joseph, 2001; Georgiadis et al., 2003;
216 Sanogo et al., 2013). However, informative priors are needed for at least four of the
217 parameters of the model, which are two sensitivities, two specificities, two covariances and
218 the true prevalence.
219
220 **Bayesians *versus* frequentist methods**

221 Previously difficult to apply because of major mathematical and computational
222 requirements, application of Bayesian approaches was facilitated by the Markov Chain Monte
223 Carlo (MCMC) methods and the availability of high quality statistical software packages
224 including JAGS (Plummer, 2003), WinBUGS (Lunn et al., 2000) and OpenBUGS (Lunn et
225 al., 2009). These approaches are now the tools of choice in many areas of application and
226 appear to offer some practical advantages over their frequentist counterparts (Greiner and
227 Gardner, 2000; Dunson, 2001; O'Hagan, 2004).
228
229 By giving the possibility to combine additional knowledge and the likelihood of
230 parameters in the population of interest in the same model, estimation is facilitated in
231 Bayesian methods. Thus, uncertainty on Se and/or Sp of the reference test, expressed as
232 probability distributions, are combined with factual observed field data to produce posterior
233 probability distributions of true prevalence and test performance (Speybroeck et al, 2012b).
234 Compared to frequentist methods, Bayesian methods also seem to offer more options and
235 flexibility to get the best possible estimates of parameters in various realistic settings.
236 Especially, the presence of conditional dependence when two imperfect tests are used can be
237 addressed in a Bayesian framework by running both models with conditional independence
238 between tests given true disease status and those with conditional dependence and checking
239 the robustness of parameters or using model selection criteria such as the Deviance
240 Information Criterion (DIC) (Berkvens et al., 2006; Dendukuri et al., 2010). Robustness of
241 estimates should also be systematically checked across a range of plausible values based on
242 the evidence to date (Enøe et al., 2000, Speybroeck et al., 2012a).
243
244 A systematic review and/or quantitative reviews summarising data using appropriate
245 meta-analytic methodologies should be particularly preferred to get informative priors on

246 diagnostic test performance (Irwig et al., 1995; Dohoo et al., 2003, EFSA, 2009). In any case,
247 application of evidence-based medicine and the quality assurance of the process to get prior
248 information are important to assess the quality of the approach. In the case of emerging
249 infectious diseases, where prior information may not be available yet and constitutes a
250 limiting factor, non-informative priors might be used. When no informative prior knowledge
251 is included in the estimation process, results of frequentist and Bayesian analysis are
252 extremely similar (Enøe et al., 2000; Dendukuri and Joseph, 2001). Whatever the priors, a
253 sensitivity analysis of prior information should be undertaken to assess its potential influence
254 on estimates (Menten et al., 2008; Sanogo et al., 2013). Thus special care should be given to
255 the selection of available information in order to get unbiased estimates (Spiegelhalter et al.,
256 2002; Berkvens et al., 2006). The procedure for incorporating available knowledge or the
257 prior information into the model and the mathematical issues have been described previously
258 (Enøe et al., 2000; Gardner et al., 2000; Dendukuri and Joseph, 2001; Branscum et al., 2005;
259 Berkvens et al. 2006).
260
261 Besides the challenges related to misclassification bias, the representativeness of data
262 regarding the population of interest and the quality assurance of the process (traceability) are
263 two key issues to be considered in both approaches. Thus, different stages of the disease and
264 age of animals should be considered and an appropriate sampling strategy should be used to
265 compose the reference population and consequently minimise sampling error and biased
266 posterior estimates. Consequences of using imperfect tests should be accounted for at the
267 analysis stage as well as the planning stage of the estimation process.
268
269 **Conclusions**

Controversies between the two approaches are more a philosophical issue than a practical issue. Although they originate from different statistical philosophies, Bayesian and frequentist approaches are two methodological options to deal with test performance and true prevalence estimation issues. While frequentists concentrate only on likelihood-based estimation, Bayesians use the likelihood and prior information for estimation. Both approaches proposed solutions to address challenges related to estimation of test performance and true prevalence taking into account field conditions. Whatever the approach, one should ensure that appropriate assumptions related to the application of a given approach hold.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Acknowledgements

The authors would like to thank the Institute of Tropical Medicine of Antwerp and the University of Liège, Belgium, for academic support.

References

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.

Berkvens, D.L., Speybroeck, N., Lesaffre, E., Saegerman, C., 2003. Comments on "Pouillot, R., Gerbier, G., Gardner, I.A., 2002. "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard. Preventive Veterinary Medicine 53, 67-81". Preventive Veterinary Medicine 59, 181-183.

Black, M.A., Craig, B.A., 2002. Estimating disease prevalence in the absence of a gold standard. Statistics in Medicine 21, 2653-2669.

Bland, J.M., Altman, D.G., 2002. Statistics notes: Validating scales and indexes. British Medical Journal 324, 606-607.

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. Preventive Veterinary Medicine 68, 145-163.

Cowling, D.W., Gardner, I.A., Johnson, W.O., 1999. Comparison of methods for estimating individual-level prevalence based on pooled samples. Preventive Veterinary Medicine 39, 211-225.

Dendukuri, N., Bélisle, P., Joseph, L., 2010. Bayesian sample size for diagnostic test studies in the absence of a gold standard: Comparing identifiable with non-identifiable models. Statistics in Medicine 29, 2688-2697.

Dendukuri, N., Joseph, L., 2001. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57, 158-167.

Dohoo, I., Martin, W., Stryhn, H., 2003. Veterinary Epidemiologic Research. AVC, Charlottetown, Prince Edward Island, Canada, pp. 85-120.

Dunson, D.B., 2001. Commentary: Practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology 12, 1222-1226.

EnØe, C., Georgiadis, M.P., Johnson, W.O., 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Preventive Veterinary Medicine 45, 61-81.

European Food Safety Authority (EFSA), 2009. Scientific Opinion of the Panel on Animal Health and Welfare (Question No EFSA-Q-2008-665), adopted on 5 June 2009. Porcine brucellosis (*Brucella suis*). EFSA Journal 1144, 1-112.

Gardner, I.A., 2002. The utility of Bayes theorem and Bayesian inference in veterinary clinical practice and research. Australian Veterinary Journal 80, 758-761.

Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Preventive Veterinary Medicine 45, 107-122.

Gart, J.J., Buck, A.A., 1966. Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. American Journal of Epidemiology 83, 593-602.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian data analysis. Chapman & Hall/CRC, London, UK, 696 pp.

Georgiadis, M. P., Johnson, W. O., Gardner, I. A. and R. Singh, R., 2003. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. Journal of the Royal Statistical Society Series C 52, 63-76.

Godfroid, J., Nielsen, K., Saegerman, C., 2010. Diagnosis of brucellosis in livestock and wildlife. Croatian Medical Journal 51, 296-305.

Greiner, M., Gardner, I.A., 2000. Application of diagnostic tests in veterinary epidemiologic studies. Preventive Veterinary Medicine 45, 43-59.

Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36, 167-171.

Hui, S.L., Zhou, X.H., 1998. Evaluation of diagnostic tests without gold standards. Statistical Methods in Medical Research 7, 354-370.

Ihorst, G., Forster, J., Petersen, G., Werchau, H., Rohwedder, A., Schumacher, M., 2007. The use of imperfect diagnostic tests had an impact on prevalence estimation. Journal of Clinical Epidemiology 60, 902-910.

Irwig, L., Macaskill, P., Glasziou, P., Fahey, M., 1995. Meta-analytic methods for diagnostic test accuracy. Journal of Clinical Epidemiology 48, 119-130.

Johnson, W.O., Gastwirth, J.L., Pearson, L.M., 2001.Screening without a ‘‘gold standard’’: the Hui-Walter paradigm revisited. American Journal of Epidemiology 153, 921-924.

Jones, G., Johnson, W.O., Hanson, T.E., Christensen, R., 2010. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. Biometrics 66, 855-863.

Joseph, L., Gyorkos, T., Coupal, L., 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. American Journal of Epidemiology 141, 263-272.

Lewis, F.I., Torgerson, P.R., 2012. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. Emerging Themes in Epidemiology 9, 9.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS-A Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing10, 325–337

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: evolution, critique and future directions (with discussion). Statistics in Medicine 28, 3049–3082

Menten, J., Boelaert, M., Lesaffre, E., 2008. Bayesian latent class models with conditionally dependent diagnostic tests: A case study. Statistics in Medicine 27, 4469-4488.

Messam, L.L.M., Branscum, A.J., Collins, M.T., Gardner, I.A., 2008. Frequentist and Bayesian approaches to prevalence estimation using examples from Johne‘s disease. Animal Health Research Reviews/Conference of Research Workers in Animal Diseases, 9, 1-23.

Meyer, N., Vinzio, S., Goichot, B., 2009: La statistique Bayésienne: Une approche des statistiques adaptée la clinique. La Revue de Médecine Interne 30, 242-249.

401 Nielsen, K., 2002. Diagnosis of brucellosis by serology. Veterinary Microbiology 90, 447-
402 459.
403
404 O'Hagan, A., 2004. Bayesian statistics: Principles and benefits. Wageningen UR Frontis
405 Series 3, 31-45. http://library.wur.nl/frontis/bayes/03_o_hagan.pdf (accessed 12
406 August 2013).
407
408 Pepe, M.S., Janes, H., 2007. Insights into latent class analysis of diagnostic test performance.
409 Biostatistics 8, 474-484.
410
411 Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs
412 sampling. URL http://www.ci.tuwien.ac.at/Conferences/DSC-
413 2003/Drafts/Plummer.pdf
414
415 Pouillot, R., Gerbier, G., Gardner, I. A., 2002. "TAGS", a program for the evaluation of test
416 accuracy in the absence of a gold standard. Preventive Veterinary Medicine 53, 67-
417 81.
418
419 Pouillot, R., 2003. Response from R. Pouillot, G. Gerbier and I.A. Gardner to D.L. Berkvens
420 et al. (2002). Preventive Veterinary Medicine 59, 185-187.
421
422 Praet, N., Dorny, P., Saegerman, C., Marcotty, T., Berkvens, D., 2006. Estimation de la
423 prévalence d'une maladie et des caractéristiques des tests diagnostiques par une
424 approche Bayésienne. Epidémiologie et Santé Animale 49, 113-130.
425
426 Qu, Y., Hagdu, A., 1998. A model for evaluating the sensitivity and specificity for correlated
427 diagnostic tests in efficacy studies with an imperfect reference test. Journal of the
428 American Statistical Association 93, 920-928.
429
430 Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test.
431 American Journal of Epidemiology 107, 71-76.
432
433 Rutjes, A., Reitsma, J., Coomarasamy, A., Khan, K., Bossuyt, P., 2007. Evaluation of
434 diagnostic tests when there is no gold standard. A review of methods. Health
435 Technology Assessment 11, 1-51.
436
437 Saegerman, C., Berkvens, D., Godfroid, J., Walravens, K., 2010. Bovine brucellosis.
438 Lavoisier & Commonwealth Agricultural Bureau – International (Ed.). In: Infectious
439 and Parasitic Diseases of Livestock, Paris, France, pp. 971-1001.
440
441 Saegerman, C., De Waele, L., Gilson, D., Godfroid, J., Thiange, P., Michel, P., Limbourg, B.,
442 Vo, T.K.O., Limet, J., Letesson, J.J., Berkvens, D., 2004. Evaluation of three serum
443 i-ELISAs using monoclonal antibodies and protein G as peroxidase conjugate for the
444 diagnosis of bovine brucellosis. Veterinary Microbiology 100, 91-105.
445
446 Saegerman, C., Vo, T.K.O., De Waele, L., Gilson, D., Bastin, A., Dubray, G., Flanagan, P.,
447 Limet, J.N., Letesson, J.J., Godfroid, J., 1999. Diagnosis of bovine brucellosis by
448 skin test: Conditions for the test and evaluation of its performance. Veterinary
449 Record 145, 214-218.
450

Sanogo, M., Thys, E., Achi, Y.L., Fretin, D., Michel, P., Abatih, E., Berkvens, D., Saegerman, C., 2013. Bayesian estimation of true prevalence, sensitivity and specificity of Rose Bengal test and indirect ELISA for the diagnosis of bovine brucellosis. The Veterinary Journal 195, 114-120.

Speybroeck, N., Devleesschauwer, B., Joseph, L., Berkvens, D., 2012a. Misclassification errors in prevalence estimation: Bayesian handling with care. International Journal of Public Health. 58, 791-795.

Speybroeck, N., Williams, C.J., Lafia, K.B., Devleesschauwer, B., Berkvens, D., 2012b. Estimating the prevalence of infections in vector populations using pools of samples. Medical and Veterinary Entomology 26, 361-371.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B. Statistical Methodology 64, 583-616.

Staquet, M., Rozencweig, M., Lee, Y.J., Muggia, F.M., 1981. Methodology for the assessment of new dichotomous diagnostic tests. Journal of Chronic Diseases 34, 599-610.

Thrusfield, M., 2005. Veterinary Epidemiology. Blackwell Science, Oxford, UK, pp. 303-330.

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. Preventive Veterinary Medicine 68, 19-33.

Vallverdú, J., 2008. The false dilemma: Bayesian vs. Frequentist. E-LOGOS Electronic Journal for philosophy 1-17.

Walter, S.D., Irwig, L.M., 1988. Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. Journal of Clinical Epidemiology 41, 923-927.

World Animal Health Organisation (OIE), 2009. Chapter 2.4.3. Bovine brucellosis. In: Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, Paris, France, pp. 1-35. http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.04.03_BOVINE_BRUCELL.pdf (accessed 12 August 2013).