Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Preventive Veterinary Medicine 68 (2005) 145–163
PREVENTIVE VETERINARY MEDICINE
www.elsevier.com/locate/prevetmed

Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling

A.J. Branscum [a], I.A. Gardner [b,*], W.O. Johnson [a]

[a] *Department of Statistics, University of California, One Shields Ave, Davis, CA 95616, USA*
[b] *Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA 95616, USA*

Received 12 June 2003; received in revised form 19 November 2004; accepted 27 December 2004

Abstract

We review recent Bayesian approaches to estimation (based on cross-sectional sampling designs) of the sensitivity and specificity of one or more diagnostic tests. Our primary goal is to provide veterinary researchers with a concise presentation of the computational aspects involved in using the Bayesian framework for test evaluation. We consider estimation of diagnostic-test sensitivity and specificity in the following settings: (i) one test in one population, (ii) two conditionally independent tests in two or more populations, (iii) two correlated tests in two or more populations, and (iv) three tests in two or more populations, where two tests are correlated but jointly independent of the third test. For each scenario, we describe a Bayesian model that incorporates parameters of interest. The WinBUGS code used to fit each model, which is available at http://www.epi.ucdavis.edu/diagnostictests/, can be altered readily to conform to different data.

Keywords: Bayesian modeling; Diagnostic tests; Sensitivity; Specificity; WinBUGS

1. Introduction

Use of Bayesian modeling of screening test accuracy has increased recently in veterinary medicine (Gardner, 2002). Diagnostic-test evaluation is particularly suited to the Bayesian framework because prior scientific information about the sensitivities and

* Corresponding author. Tel.: +1 530 752 6992; fax: +1 530 752 0414.
E-mail address: iagardner@ucdavis.edu (I.A. Gardner).

0167-5877/$ – see front matter
doi:10.1016/j.prevetmed.2004.12.005

specificities of the tests and prior information about the prevalences of the sampled populations can be incorporated. Because the review article of Enøe et al. (2000) described many frequentist and Bayesian diagnostic-test modeling strategies, we focused on the computational aspects involved in a fully Bayesian analysis to estimate parameters of interest from these models. We used the freeware program WinBUGS (Spiegelhalter et al., 1996) for all modeling. Readers unfamiliar with this software are referred to the WinBUGS online tutorial at the website http://www.statslab.cam.ac.uk/~krice/winbugsthemovie.html. This tutorial provides a step-by-step introduction to using WinBUGS for Bayesian statistical modeling.

In general, the Bayesian approach to inference about a generic parameter θ combines prior information about θ (through the prior $p(\theta)$) with the data (through the likelihood $p(data|\theta)$) to obtain the posterior distribution of θ, $p(\theta|data)$. Then, one can use the mean, median, or mode of this posterior distribution as an estimate of θ. Often, the posterior distribution of a parameter does not belong to a well-known family of distributions (such as the normal, beta, or gamma families). However, if one can obtain a sample from $p(\theta|data)$, then a Monte Carlo based estimate of θ can be calculated. For instance, the mean of, say, N independently simulated values from the posterior distribution of θ would be a Bayesian estimate of θ. The lower and upper endpoints of a 95% probability interval for θ are the 2.5th and 97.5th percentiles of the Monte Carlo sample. We note that we prefer the phrase "probability interval" over the commonly used "credibility interval" because "probability interval" is more descriptive and more closely related to the interpretation of the interval.

Sampling from posterior distributions can, in many cases, be accomplished by using the WinBUGS package. In WinBUGS, only the likelihood and the prior distributions need to be specified by the user. The posterior distributions of the model parameters then are sampled automatically (iteratively sampling from the so-called "full-conditional" distributions). If the full-conditionals belong to a recognizable family of distributions, they are sampled exactly. Failing in this (and if the full-conditional is log-concave) the method of adaptive rejection sampling (Gilks and Wild, 1992) is used. Failing in this, a Metropolis algorithm (Robert and Casella, 1999) that involves an adaptive phase is used.

We review Bayesian models designed to evaluate diagnostic-test accuracy that were presented in Enøe et al. (2000); Johnson et al. (2001); Dendukuri and Joseph (2001) and Georgiadis et al. (2003). Each model is fitted to previously published data using WinBUGS. We also describe a new scenario in which the diagnostic-test accuracy of two conditionally dependent tests can be estimated given the availability of an independent third test. This third test may be essentially a gold-standard or reference test. The models we present generalize those of Enøe et al. (2000); Johnson et al. (2001) and Georgiadis et al. (2003) in that the prevalences of the sampled populations are modeled using a mixture distribution that allows for zero infection prevalence with non-zero probability.

2. Background

Consider the following scenario. Assume that $L \geq 1$ populations are sampled independently and n_k animals are selected randomly from population k. Each sampled