Determining the Infection Status of a Herd

Timothy E. HANSON, Wesley O. JOHNSON,
Ian A. GARDNER, and Marios P. GEORGIADIS

This article presents hierarchical models for determining infection status and prevalence of infection within a herd given a hypergeometric or binomial sample of animals that have been screened with an imperfect test. Expert prior information on the infection status of the herd, diagnostic test accuracy, and herd prevalence is incorporated into the model. Posterior probabilities versus prior probabilities of infection are presented in the novel form of a curve, summarizing the probability of infection over a range of possible prior probability values. We demonstrate the model with serologic data for *Mycobacterium paratuberculosis* (Johne's disease) in dairy herds.

Key Words: Bayesian approach; Gibbs sampling; Prevalence; Screening test; Sensitivity; Specificity.

1. INTRODUCTION

Local and regional trade in animals is in part dependent on the validity of assurances that a herd of animals is "infection-free," or has a prevalence of infection less than a low threshold value. This classification is important to herd owners because it provides a competitive marketing advantage for their livestock. Alternatively, if the presence of a pathogen is detected early, and at low prevalence, owners may treat and expediently eradicate the pathogen, conserving resources and preventing a potentially costly, more widespread infection.

In the United States, herd infection-status certification has received increased attention due to the new Voluntary Johne's Disease Herd Status Program (VJDHSP) for cattle (Bulaga 1998), whose goal is to minimize the risk of transmission of *Mycobacterium paratuberculosis* among herds. Issues associated with the determination of herd prevalence are also important to voluntary programs such as VJDHSP as well as to the monitoring of federally regulated animal diseases such as brucellosis and pseudorabies.

Timothy E. Hanson is Assistant Professor, Department of Mathematics and Statistics, University of New Mexico, Albuquerque 87131 (E-mail: hanson@math.unm.edu). Wesley O. Johnson is Professor, Department of Statistics, University of California, Davis 95616. Ian A. Gardner is Professor, Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA 95616. Marios P. Georgiadis is Epidemiologic Research Associate, Laboratory of Clinical Bacteriology, Parasitology, Zoonoses and Geographical Medicine, Faculty of Medicine, University of Crete, Greece.

Consider the following scenario: a herd of animals is to be classified as infected or noninfected based on the prior expert-elicited probability that the herd is infected and a random sample of screened animals from the herd. The reference test will typically be imperfect and we model the test accuracy a priori based on the kit manufacturer's leaflet, published results, and/or expert opinion. The resulting inference of interest is the posterior probability, given the results of the screened animals and the expert opinion, that the herd is infected. Additionally, we may obtain an estimate of the number (or prevalence) of infected animals in an infected herd and refined estimates of herd-specific test accuracy.

We allow the prevalence to be zero with positive probability and thus these models are well-suited for the classification of a herd's infection status. In order to specify the prior probability of infection, a herd veterinarian, government official, or other expert may consider the proportion of infected herds, p_I, among those that are similar to the one under consideration, for example, exhibiting signs consistent with the presence of the infectious agent and with similar management practices. This is operationally *identical* to simply providing a prior probability of infection, also p_I, for the specific herd under consideration, when no herd-specific information is available. On the other hand, if herd-specific information is available, it would be considered as well in the determination of a value for p_I, then regarded as the prior probability of infection for the specific herd in question. The posterior probability of herd infection is readily calculated for a given sample of tested animals from the herd. A plot of posterior infection probabilities versus corresponding prior probabilities is a convenient graphical summary of the herd's status that may be used by the veterinarian to determine the likelihood of herd infection over the range of prior probabilities. Alternatively, we may model the proportion of similar herds that are infected, p_I, with a beta distribution. We examine both of these approaches.

The models we examine are over-parameterized and thus a strictly frequentist approach is impossible without unrealistic or unusual restrictions on the model parameters. Moreover, in models with small sample sizes, inference is based on asymptotic normal approximations that might be highly inaccurate. The Bayesian approach avoids the nonidentifiability problem by placing prior distributions on all model parameters and thus resulting inferences will necessarily be dependent on these prior distributions. Existing models (e.g., Cameron and Baldock 1998a,b; Cannon 2001) *do not* allow for uncertainty in the accuracy estimates, which may be problematic for a number of reasons (Greiner and Gardner 2000) that are subsequently discussed. We also note that a reliable range of test accuracy values is often available, obtained from published reports or from subject-matter experts.

In the models we consider, full conditional distributions of model parameters are readily found, and hence inference is easily obtained by setting up an iterative "Gibbs sampler" algorithm. We do not discuss the specifics of using Markov chain Monte Carlo (MCMC) methods for posterior inferences, but instead refer the reader to classic articles such as those by Gelfand and Smith (1990) and Tierney (1994).

2. THE STATISTICAL MODELS

We examine two models that provide posterior probabilities that a herd is infected, given a random sample of tested animals. The first model assumes that animals are drawn

from a finite population; the second model assumes the herd is infinite, but provides a very good approximation to the first model when the herd size is large relative to the number sampled.

These models provide a posterior distribution on the proportion of infected animals and therefore may be used with a "threshold" value, $\pi \leq \pi_T$, for declaring a herd noninfected (Cannon and Roe 1982; Cameron and Baldock 1998a,b). The models we consider easily provide the posterior probability of this event.

2.1 HYPERGEOMETRIC SAMPLING

We examine the problem of assigning a probability to the event that a herd of animals of size N has at least one infected member, given a random sample of size n from the herd that has been evaluated with an imperfect screening test. We denote an animal's infection state by $D = 0, 1$ for absence/presence of infection and the test results as $T = 0, 1$ for negative/positive result. We observe x out of n testing positive. Of interest is the posterior distribution of k, the number of infected animals in the entire herd, and in particular the probability that at least one animal is infected, namely $P(k \geq 1|x)$.

The prior probability that the herd is infected is denoted p_I and is provided by an expert. This quantity may simply be an estimate of the proportion of infected herds of similar size and having similar health management practices in an area. We may examine the posterior probability of infection as a function of this prior probability $P(k \geq 1|x, p_I)$ for different values of p_I. It is also possible to model p_I with a beta prior to reflect a subjective plausible *range* of infection probability, discussed later in this section. Knowledge of a diagnostic test's sensitivity $\eta \equiv P(T = 1|D = 1)$ and specificity $\theta \equiv P(T = 0|D = 0)$ is modeled using independent beta priors

$$\eta \sim \text{beta}(a_\eta, b_\eta) \perp \theta \sim \text{beta}(a_\theta, b_\theta); \tag{2.1}$$

see Johnson and Gastwirth (1991) and Johnson, Gastwirth, and Pearson (2001). We note that the models we consider with sensitivity and specificity *known*, that is, set at values η_0 and θ_0 with probability one, are identifiable and maximum likelihood estimators exist for all remaining parameters. Therefore, accurate prior information on test accuracies will yield accurate posterior estimates of the probability of infection. Some approaches simply fix the test accuracies at published point estimates and proceed. However, by not modeling the uncertainty in the test, posterior credible intervals will be artificially narrowed. Estimates of sensitivity and specificity may be uncertain due to limited sample sizes in the test evaluation studies and estimates may vary with factors such as animal age, disease stage, and geographic location (Greiner and Gardner 2000). A limitation of the current VJDHSP guidelines is that sensitivity and specificity are assumed to be known *exactly*, with no uncertainty.

A discretized beta(a_π, b_π) distribution is placed on the distribution of the number infected k out of N, given that k is at least one:

$$p(k|k \geq 1) \propto k^{a_\pi - 1}(N - k)^{b_\pi - 1}. \tag{2.2}$$

The beta distribution is the standard prior for a proportion and, as long as the expert's opinion can be expressed in terms of a unimodal density, it should be perfectly adequate. Otherwise, a finite mixture of beta distributions could be used.

The distribution of k is necessarily

$$p(k) = \left\{ \begin{array}{cc} 1 - p_I & k = 0 \\ p(k|k \geq 1)p_I & 1 \leq k \leq N \end{array} \right\}. \tag{2.3}$$

We introduce latent data to facilitate the implementation of a Gibbs sampler. The number of animals that are truly infected is denoted z and of those the number that are correctly classified as infected is denoted v. We have the complete 2×2 table as follows:

	$D = 1$	$D = 0$	
$T = 1$	v	$x - v$	x
$T = 0$	$z - v$	$n - x - z + v$	$n - x$
	z	$n - z$	n

The distribution of the number sampled that are infected, z, given the number k out of N that are infected, is hypergeometric

$$z|k \sim \text{hypergeometric}(k, N, n). \tag{2.4}$$

We assume that the test T acts independently within diseased and nondiseased populations and hence:

$$\begin{array}{rcl} v|z, \eta, \theta, k & \sim & \text{binomial}(z, \eta), \\ x - v|z, \eta, \theta, k & \sim & \text{binomial}(n - z, 1 - \theta). \end{array} \tag{2.5}$$

The joint distribution $p(v, x - v, z|\eta, \theta, k)$ is derived in Appendix A, where the full conditionals necessary for Gibbs sampling are also presented. In theory, the posterior distribution for $k|x$ may be obtained by marginalizing the parameters η and θ in the joint posterior $p(v, x-v, z|\eta, \theta, k)P(\eta, \theta, k)$ and summing over v and z, but this sum approaches 2^n terms and is impractical to calculate for all but very small sample sizes. Approximations have been suggested in the literature for the non-Bayesian model, with sensitivity and specificity fixed (Cameron and Baldock 1998a,b; Cannon 2001), but these fail for the models we consider. We have found Gibbs sampling to be satisfactory in practice, even with very large herds, say $N > 1,000$.

In order to calculate the posterior probabilities of infection over a grid, it suffices to calculate the posterior probability for a single value of p_I since the posterior odds of infection, $P(k > 0|x, p_I)/\{1 - P(k > 0|x, p_I)\}$, divided by the prior odds, $p_I/(1 - p_I)$, is simply the Bayes factor (BF) for comparing the hypothesis of positive prevalence to no infection, which is free from p_I. It is then possible to use the BF to calculate the posterior probabilities of infection according to the formula $\frac{p_I}{1-p_I}\text{BF}/[1 + \frac{p_I}{1-p_I}\text{BF}]$. More discussion of Bayes factors for this problem is given in Section 3.1.

2.2   Binomial Sampling

If N is large and n is small relative to N, we can accurately approximate the above model by assuming binomial sampling of the herd. We assume the herd size is infinite and therefore do not consider the number of infected animals k, but instead model the prior infection prevalence $\pi = P(D = 1)$ as a mixture of a continuous distribution on $(0, 1)$ and point mass at zero $\delta_{\{0\}}$

$$\pi \sim p_I \text{beta}(a_\pi, b_\pi) + (1 - p_I)\delta_{\{0\}}. \tag{2.6}$$

This approach allows us to calculate probabilities such as $P(\pi > 0|x)$ and $P(\pi < \pi_T|x)$ for example. As in the first model we detect x out of n animals that test positive and therefore the data x given the other parameters are distributed binomial

$$x|\eta, \theta, \pi \sim \text{binomial}\{n, \pi\eta + (1 - \pi)(1 - \theta)\}. \tag{2.7}$$

When N is even moderately large, this simplification increases the efficiency of the MCMC algorithm enormously and provides very accurate approximations, as we illustrate in the next section.

The full conditional distributions of the test accuracies and the latent data are derived and presented in Appendix B. Bayes factors are easily calculated.

Remark 1.   *We note here that our Bayesian model for binomial data is not particularly novel though the application of it to our particular risk analysis problem is. With perfect sensitivity and specificity, and with $p_I = 1$, our model reduces to the standard binomial model with a conjugate beta prior (Casella and Berger 2002, pp. 324–325). This particular model is somewhat ubiquitous in the Bayesian literature. Moreover, the model without perfect sensitivity and specificity, but $p_I = 1$, was also considered in Gastwirth, Johnson, and Reneau (1991) in the context of screening for HIV infection, and in the subsequent work of Joseph, Gyorkos, and Coupal (1995) and Johnson, Gastwirth, and Pearson (2001) in the context of screening without a gold standard based on two tests. There are numerous other references to relevant papers given in Enøe, Georgiadis, and Johnson (2000). However, the contextual application of the presented model is novel to our knowledge.*

2.3   A Prior on the Proportion of Infected Herds

Instead of specifying a particular value for p_I, and in particular when p_I is regarded as the proportion of infected herds in the population, it may be desirable to provide a distribution for it. This is easily incorporated into the model as

$$p_I \sim \text{beta}(a_p, b_p). \tag{2.8}$$

The full conditionals for both models are readily calculated. For the first model that assumes a finite number of animals, N, the full conditional for p_I is written:

$$p_I|\text{else} \sim \text{beta}(a_p + 1 - I_{\{0\}}(k), b_p + I_{\{0\}}(k)).$$

For the second model where herd size is assumed to be infinite:

$$p_I | \text{else} \sim \text{beta}(a_p + 1 - I_{\{0\}}(\pi), b_p + I_{\{0\}}(\pi)),$$

where $I_A(x)$ is equal to one when $x \in A$ and zero otherwise. We note that it is then straightforward to model the joint probability of infection for several herds simultaneously and to model the herd-to-herd prevalences of disease for infected herds with a beta hyperprior or a more general mixture of Dirichlet processes centered around a beta hyperprior (Hanson, Johnson, and Gardner 2003; Suess, Gardner, and Johnson 2002).

Remark 2.   *A referee has pointed out the importance of modeling prior information* very carefully *for this problem. This is especially clear because the data involve only a simple binomial or hypergeometric count for the number of positive test outcomes, while final inferences depend on three unknown parameters, the prevalence, and two accuracies. It would be easy for two experts to disagree sufficiently that their posterior inferences could be appreciably different. Thus, a sensitivity analysis of the final conclusions to perturbations in the choice of prior may be warranted. However, the main application we envision is that of a herd owner, in collaboration with the herd veterinarian, attempting to assess the disease status of one of his/her herds. Based on their experience in dealing with the given herd, they will be the expert who provides prior input to be combined with the binomial/hypergeometric count of test positives. The ultimate conclusion of the analysis is thus a coherent combination of the two forms of information resulting in* the herd owner's personal posterior probability of herd infection. *Then she/he must ultimately decide—based on either formally, as in Geisser and Johnson (1992), or informally taking account of the costs associated with making Type I and Type II errors—whether to perform some kind of intervention, or not.*

Alternatively, it may be a governmental veterinarian who would be asked to provide prior input. In this instance, the input may not be specific to the particular herd but rather on p_I regarded as the prevalence of similar herds that are infected. If a particular herd is assessed to be infected based on the combined input, the government may force an intervention such as quarantine of the herd, for example, or in the case of Johne's disease, inclusion or exclusion from one of the various categories of the VJDSHP. If there are multiple governmental veterinarians with differing opinions about test accuracy and herd prevalence, the corresponding posteriors will likely be different. But we hope that, over time, sufficient information will become available for them to agree on the prior inputs. Failing in that, and in those instances where different posteriors are obtained, some form of arbitration involving the relative qualifications of the veterinarians and/or the relative costs of errors may be necessary to make decisions about individual herds.

3. ILLUSTRATIVE EXAMPLES

We demonstrate our models on one simulated and several real datasets, including two Colorado and four California cattle herds suspected of being infected with *Mycobacterium paratuberculosis*, the causative agent of Johne's disease.
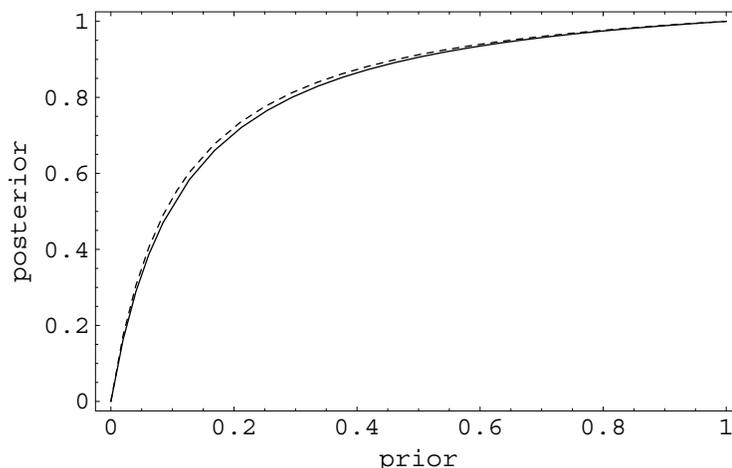
Figure 1. Posterior versus prior probability of herd infection: solid is hypergeometric model, dashed is binomial.

3.1 SIMULATED DATA

Assume that the true proportion of infected animals is $\pi = k/N = 0.2$. We construct a dataset in the following way. Assume the sensitivity $\eta = 0.7$ and the specificity $\theta = 0.9$. Then out of $n = 100$ animals we expect to have 20 infected and therefore we expect x to be

$$E(x) = 20P(T = 1|D = 1) + 80P(T = 1|D = 0) = 20(0.7) + 80(0.1) = 22.$$

Assume that the total herd size is $N = 500$. We set the mean of the sensitivity and specificity priors at their true values: $\eta \sim \text{beta}(70, 30)$ and $\theta \sim \text{beta}(90, 10)$. The data are $x = 22$, $n = 100$, and $N = 500$. We fit both the hypergeometric model, (2.1)–(2.5), and the binomial approximation, (2.1), (2.6), and (2.7). A plot of the posterior versus the prior probability of infection is shown in Figure 1. The models agree quite well even though the herd size is only five times the number sampled. Notice that the posterior probability of infection is greater than 0.5 when the prior probability is larger than 0.1; we have strong evidence of infection in this example.

Because

$$P(\pi > 0|x) = \int_0^1 P(\pi > 0|p_I, x)p(p_I|x)dp_I,$$

it is interesting to note that, when $p_I|x \sim U(0, 1)$—that is, all proportions of similar herds that are infected are equally likely *regardless of the data*—the area under the curve is then the posterior probability of infection. This area is given as a function of the Bayes factor (BF) by

$$\int_0^1 P(\pi > 0|p_I, x)dp_I = \text{BF}(\text{BF} - \log(\text{BF}) - 1)/(\text{BF} - 1)^2.$$

For the simulated data, this area is 0.83 using either model. The shape of the curve—that is, whether it is concave or convex—gives an indication of the infection status of the herd: infected or noninfected, respectively, with the strength of the concavity or convexity being related to the strength of the infection determination. Moreover, the Bayes factor for testing that the prevalence is nonzero versus zero is the ratio of marginal probabilities for x under these hypotheses, namely BF $= p(x|\pi > 0)/p(x|\pi = 0)$, and the curve defined by the plot of $P(\pi > 0|x, p_I)$ versus p_I will be a line with intercept zero and slope one if BF $= 1$, it will be strictly concave as in Figure 1 if BF > 1 and it will be strictly convex if BF < 1.

Convergence of the Gibbs sampler is conservatively assessed by running several parallel chains from different starting values. We check that infection probability estimates from the parallel chains agree to at least two decimal places and graphically check convergence using running quantile plots. Additionally, one may run a separate Gibbs sampler for each value p_I on a grid of values, say $p_I \in \{i/20 : i = 1, \ldots, 19\}$. As discussed earlier, a plot of the posterior probability of infection versus these values should be a smooth curve, providing another check that the number of iterates taken in the Gibbs samplers is adequate. We have found the number of Gibbs sampling iterates conservatively needed to range from 10,000 to 10,000,000 depending on the model, data, and prior specification.

We find that the binomial model code typically runs about 100 times faster than the hypergeometric code (e.g., 15 seconds for the binomial model versus half an hour for the hypergeometric model in FORTRAN 90 on a 933 Mhz Pentium III with one gigabyte of RAM). FORTRAN 90 programs for both models are available for download at http: //www.epi.ucdavis.edu/diagnostictests/.

3.2   TWO COLORADO DAIRY HERDS

We consider the classification of two Colorado dairy herds, labeled B and J, with respect to being infected with *M. paratuberculosis*. A prior distribution on the proportion of similar herds that are infected, p_I, as well as herd-level prevalence priors given that the herd is infected, $\pi^* = k/N|k \geq 1$, were provided by Dr. Heather Hirst of Colorado State University, Fort Collins, CO, a veterinarian working with each herd. Her priors were based on management practices used in each herd that are known to affect the risk of *M. paratuberculosis* infection and clinical evaluations of cows. The diagnostic test used for all screening is the enzyme-linked immunosorbent assay (ELISA) with a preset cut-off value for the determination of animal infection.

We base inference on the hypergeometric model since each herd was completely sampled. We compute the posterior probability that each herd is infected given the data and we plot the posterior probabilities of the herd being infected $P(k > 0|x, p_I)$ against the prior infection probabilities $p_I \in (0, 1)$. Given that the herd is infected, the prevalence was modeled as a discretized beta distribution as in (2.2). These priors, as well as the priors on p_I for each herd (2.8), are presented in Table 1. We denote the qth quantile of any random variable R as R_q.

Table 1.   Priors on Probability of Infection and Prevalence: Two Colorado Dairy Herds

Herd	$p_{I, 0.5}$	$p_{I, 0.95}$	Prior	$\pi^*_{0.5}$	$\pi^*_{0.95}$	Prior
B	0.20	0.30	beta(10.79,42.19)	0.07	0.13	beta(5.35,67.05)
		$p_{I, 0.05}$			$\pi^*_{0.05}$	
J	0.98	0.90	beta(24.60,0.793)	0.25	0.05	beta(1.724, 4.547)

The sensitivity of the ELISA test for detection of serum antibodies to *M. paratuberculosis* is highly dependent on the stage of infection (stage I: preclinical, not shedding bacteria in feces; stage II: preclinical, shedding bacteria in feces; or stage III: clinical) and the number of *M. paratuberculosis* in feces (Sweeney, Whitlock, and Buckley 1995; Whitlock, Wells, Sweeney, and Van Tiem 2000). In stage II infection ELISA sensitivity is positively correlated with the number of *M. paratuberculosis* in feces (Whitlock et al. 2000).

Much conjecture exists about estimates for the sensitivity of the ELISA, especially for stage II infection, in part because of choice of the reference test. Recent estimates for ELISA sensitivity in adult cattle vary from 15% to 85% depending on the stage of infection (Dargatz et al. 2001) and whether "heavy" fecal shedders of *M. paratuberculosis* have been culled from the herd. In the latter case, ELISA sensitivity decreases with time in some herds (Whitlock et al. 2000). Estimates also depend on the choice of the decision threshold for designation of a positive or negative test result, and there are several variants of the test in different laboratories. In addition, none of the test evaluation studies used samples from Colorado or California, which have larger herd sizes and different production practices than herds in the midwest and eastern United States, which had been used in the original studies.

Because we were unable to find a single appropriate estimate for sensitivity in the published literature, we asked two subject-matter experts (Drs. Michael Collins and Robert Whitlock) who have evaluated the ELISA separately (Collins and Sockett 1993; Sweeney et al. 1995; Whitlock et al. 2000) to give their expert opinion. We asked the experts for their best guesses (modes) and the values that they were 95% sure that sensitivity of the commercially available ELISA (IDEXX Laboratories, Westbrook, ME) was below when used for testing an "average" dairy cow in stage II infection. Stage II infection includes cows shedding few, moderate, or large numbers of *M. paratuberculosis* in feces and therefore their input required a synthesis of information about the proportion of cows in each of these categories in an "average" herd and the sensitivity of each category. If it were believed that animals in all three stages were present, a more complicated synthesis would be required. For the Colorado dairy herds we used the priors provided by Michael Collins of the University of Wisconsin, $\eta \sim \text{beta}(58.82, 174.47)$ (mode = 0.25) independent of $\theta \sim \text{beta}(272.41, 6.54)$ (mode = 0.98).

Herd B introduces animals from other herds but has no clinical signs in cows suggestive of Johne's disease. The expert believes that there is a 50/50 chance that the proportion of herds like this one is either above or below 0.20, and they are 95% sure that this proportion is less than 0.30, that is, $P(p_I < 0.20) = 0.50$ and $P(p_I < 0.30) = 0.95$. Her best guess
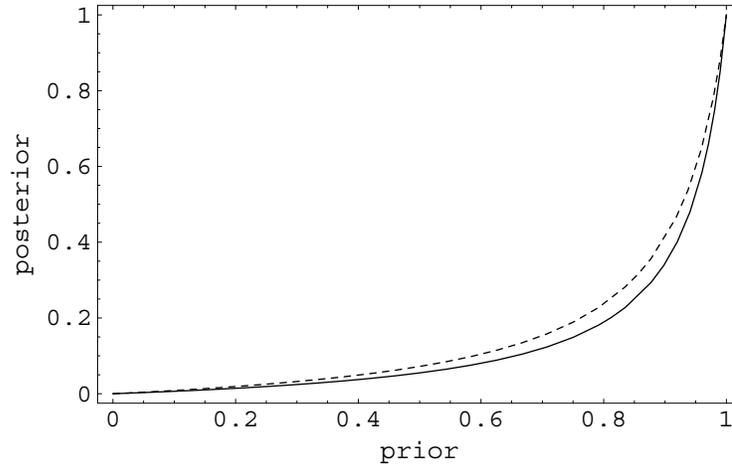
Figure 2.   Posterior versus prior probability of herd infection: herds B (dashed) and J (solid).

for the probability that this particular herd is infected would be 0.20. Of the $N = n = 493$ animals in the herd at the time of testing, $x = 6$ were ELISA positive. Figure 2 shows the posterior probability of infection as a function of the prior probability. The curve is strongly convex, indicating that herd B is likely infection-free. The Bayes factor for these data is 0.08, indicating that the data are 13 times more likely under the model that assumes the herd is infection free than under the model that assumes infection is present. When we incorporate the subject-matter expert's prior information that this herd is infected, we find that the posterior probability of infection is $P(k > 0|x) = 0.02$. Indeed the number testing positive required to place the posterior probability of infection above 0.9 is $x = 38$, much larger than the 6 that were observed.

Herd J is a typical dry-lot dairy farm, it introduces animals, and has historically had a few clinical cases of Johne's disease. The expert's best guess for the probability that this herd is infected is 0.98, and they are 95% sure that the proportion of herds like this one that are infected is at least 0.90, $P(p_I > 0.9) = 0.95$. Of the $N = n = 669$ animals in the herd at the time of testing (June 2000), $x = 14$ were positive by ELISA. The Bayes factor for these data is 0.06, indicating the data are about 17 times more likely under the model that assumes the herd is infection-free than under the model that assumes infection is present. Figure 2 displays the posterior versus prior probability that herd J is infected. We see that there is strong evidence that this herd is not infected at moderate prior infection probabilities p_I.

Because the prior estimate of the specificity is 0.98, roughly 13 false positives are expected in the data. Thus, one might expect the Bayes factor to be closer to 1 than it is. However, because the prior on π^* is focused on 0.25, the infection model for the data would predict over 25 true positives or a total of 38 or more positives. Thus, the model of no infection appears to be more plausible with the given prior input. As part of our sensitivity analysis, we decided to modify the prior on π^* to be a mode zero beta(1,b) distribution

Table 2.   Priors on Probability of Infection and Prevalence: Four California Dairy Herds

Herd	$p_{I,0.5}$	$p_{I,0.05}$	Prior	$\pi_{0.5}^*$	$\pi_{0.95}^*$	Prior
201	0.98	0.50	beta(1.976,0.228)	0.07	0.10	beta(17.42,227.29)
202	0.90	0.02	beta(0.482,0.200)	0.03	0.05	beta(8.59,267.49)
203	0.99	0.05	beta(0.483,0.118)	0.09	0.15	beta(7.85,76.34)
212	0.95	0.10	beta(0.715,0.198)	0.15	0.20	beta(24.50,137.28)

where b was selected so that the expert would be 95% sure that the nonzero prevalence was less than 0.3: beta(1,8.4). Leaving other input the same, the Bayes factor was about 5. The posterior probability of infection based on the data and the original expert opinion is $P(k > 0|x) = 0.65$, not surprising considering her strong prior belief that the herd is infected. The same probability based on the modified expert opinion is 0.27.

3.3   FOUR CALIFORNIA DAIRY HERDS

We examine four California dairy herds with herd sizes in the 500–1,000 range. From each herd $n = 60$ animals were tested by ELISA for *M. paratuberculosis*. The number sampled is considerably less than the herd sizes and we use the binomial sampling model for each herd. Priors on test accuracy were provided by two nationally known experts, those provided by Dr. Michael Collins, used in the previous example, and priors provided by Dr. Robert Whitlock of the University of Pennsylvania: $\eta \sim$ beta$(4.46, 14.84)$ (mode $=$ 0.20) independent of $\theta \sim$ beta$(384.13, 16.96)$ (mode $= 0.96$). Beta distributions for the prior probability of infection p_I for each herd, and the prevalence π^* of the herd given it is infected were provided by Randy Anderson of the California Department of Food and Agriculture, as part of a study evaluating collection of prior information from many farms in a standardized way. Dr. Anderson synthesized information, based on use of a risk factor questionnaire about management practices used by the owner to prevent transmission of *M. paratuberculosis* and his clinical evaluations of cows, into a prior for p_I and π^* for each herd. The resulting posterior information would be used by the government veterinary service for classification of herd status for *M. paratuberculosis* and estimation of the proportion of infected herds in a specific geographic region. The resulting priors are in Table 2.

The herd owners and veterinarians are interested in the posterior probability that each herd is infected and also in an upper 95% bound on the prevalence: the value $\pi_{0.95}$ such that $P(\pi \leq \pi_{0.95}|x) = 0.95$. These results are shown in Table 3, along with the posterior

Table 3. Posterior Probabilities of Infection Based on Priors From Two Different Experts: Four California Dairy Herds

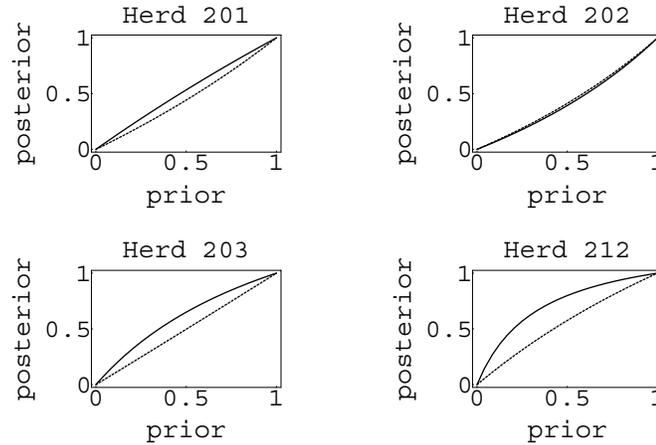| Herd	Michael Collins		Robert Whitlock		
	$P(\pi > 0 \mid x)$	$\pi_{0.95}\mid x$	$P(\pi > 0 \mid x)$	$\pi_{0.95}\mid x$	$x/60$
201	0.91	0.10	0.87	0.10	2
202	0.61	0.05	0.62	0.05	0
203	0.88	0.15	0.81	0.14	3
212	0.93	0.20	0.83	0.19	4

Figure 3.   Posterior versus prior probability of herd infection: Collins is solid, Whitlock is dashed.

probability of infection corresponding to each test accuracy expert. The posterior infection probabilities for the two experts are similar across herds; the 95th posterior prevalence percentiles are identical across experts, for each herd. There is moderate posterior evidence, mainly driven by high prior probabilities, that herds 201, 203, and 212 are infected but it is a coin toss as to whether herd 202 is infected given the existing information.

Because the posterior probability of herd infection can depend somewhat heavily on the prior guess for p_I, we also present plots for $P(\pi > 0|x, p_I)$ versus p_I for all four herds in Figure 3, where the two curves correspond to the two expert priors on test accuracy. The plots are concave for herds 201 ($x = 2$), 203 ($x = 3$), and 212 ($x = 4$), indicating that Bayes factors are greater than 1 for these herds, and convex for herd 202 ($x = 0$), indicating that the Bayes factor is less than one for this herd. The difference between the expert priors has an obvious impact on posterior inferences. For example, under the Whitlock prior, one expects roughly twice as many false positives than would be expected with the Collins prior. The effect of this is perhaps most noticeable for herd 212 where the observed $x = 4$ in conjunction with the Collins prior gives a somewhat stronger indication of herd infection than the posterior based on the Whitlock prior.

We examine how the posterior probability of infection changes with perturbations of the priors on infection status p_I and conditional prevalence π^* for herd 212. We examine uniform priors for both, and use the test accuracy prior of Michael Collins. Collins also provided prior information on the proportion of herds infected in the United States and for those herds with Johne's disease, the proportion of infected animals; the priors are $p_I \sim$ beta$(4.84, 3.56)$ and $\pi^* \sim$ beta$(1.51, 10.78)$. We may use this information for "typical" herds in lieu of Dr. Anderson's herd-specific priors in a sensitivity analysis. The results are in Table 4.

We see that in all cases, the posterior probability of infection is above 0.60 and most are

Table 4. Posterior Probabilities of Infection for Herd 212 From Various Prior Distributions (Collins prior on ELISA accuracy)

p_I prior	π^* prior	$P(\pi > 0 \mid x)$
beta(0.715,0.198)	beta(24.50,137.28)	0.93
beta(0.715,0.198)	beta(1.51,10.78)	0.92
beta(0.715,0.198)	beta(1,1)	0.85
beta(0.715,0.198)	beta(24.50,137.28)	0.93
beta(4.84,3.56)	beta(24.50,137.28)	0.83
beta(1,1)	beta(24.50,137.28)	0.78
beta(1,1)	beta(1,1)	0.61
beta(4.84,3.56)	beta(1.51,10.78)	0.80

above 0.80; even with diffuse priors on p_I and π^* we have evidence of infection. However, it is clear that the choice of prior specification does have an effect on posterior inference, and in this specific example diffuse priors tend to mitigate the posterior probability of infection towards 0.5.

3.4 DISCUSSION

In a recent followup analysis of three of the California herds, it was determined by culture testing (specificity = 1) that the two positive ELISA results for herd 201 were actually false positive, that two of the three positives for herd 203 were false, and that two of the four positives for herd 212 were also false. Moreover, herds 203 and 212 were confirmed infected based on culture results from the matched samples (blood and feces). An additional factor not taken into account in our analysis is that the serologic tests performed actually result in a numerical score. Our analysis follows tradition where a cut off value was used and a positive or negative outcome determined by whether the cut off was exceeded or not, respectively. In the followup analysis, it was noted that the ELISA positive scores that corresponded to culture negative outcomes were near the cut off while those corresponding to culture positive scores were much larger. Thus, it may be fruitful in future risk analyses to take account of the quantitative nature of the original serologic data.

Some readers may be concerned that the differences between the two experts' priors resulted in somewhat different inferences for California herd 212. However, this is merely a reflection of the well-known fact in science: experts in the same area often disagree. One of the goals of science is to ultimately provide sufficient data that, even when (separately) combined with the information from disparate sources, the combined information will result in a reconciliation of prior differences. In our situation, this means eventually observing "sufficient" data that the two experts' posteriors would be similar to one another. Part of the discrepancy here may well be due to the fact that the "synthesis" of information about test accuracy related to concentration of shedding is inherently different for the two experts.

There are other diseases for which the information about test accuracy should be less controversial. For example, when a gold standard or "near" gold standard test exists, it is possible to have accurate prior information about test accuracy, even if the test is prohibitively expensive for general screening. This is certainly the case for HIV testing (Johnson and Gastwirth 1991) where, for surveillance purposes, the gold standard test would probably be prohibitively expensive for general use in many parts of Africa, for example. In the veterinary literature, examples where a consensus on test accuracy exists include the ELISA antigen detection of heartworm in dogs (Courtney and Zeng 1993) and the modified agglutination test on serum for toxoplasmosis in pigs (Dubey et al. 1995).

In our analysis of the California herds, the number of positive results ranged from 0–4 out of 60 animals tested in each herd. For each herd, the posterior probability of herd infection was smaller than the expert's prior estimate for this probability. Even with zero positives, there is still an inclination to believe that herd 202 is "diseased" based on the posterior probability. Increasing the sample size would help to give more definitive results in the face of such strong prior beliefs.

If π^* is thought to be less than or equal to 0.5, but relatively little is known about the actual value, it may be sensible to choose a beta(1,b) prior ($b \geq 1$). However, our limited experience suggested that models are somewhat robust to prior on π^*.

A limitation of our approach is that only two infection states are modeled, while often there may be three or more verifiable infection states (e.g., absent, preclinical, and clinical). Allowing for several infection states with differing test accuracies depending on the state will yield an improved model and provide several additional herd classification scenarios. For example, if a herd is devoid of animals that are either preclinical or clinical, we may classify the herd as infection-free. If there are preclinical animals but no clinical animals, we might classify the herd differently than if clinical animals are present. This additional information may be useful for decision-making in herd management and more cost-effective animal health care. We are currently examining a model for Johne's disease in which three infection states are modeled and animals are tested over time. We are considering both the ELISA and a fecal culture test and allowing for dependence between the two tests.

4. CONCLUSIONS

We have presented hierarchical Bayesian models for determining the infection status of a herd given a random sample of animals that have been screened with an imperfect test. These models incorporate uncertainty in the screening test accuracy as well as prior information on the probability of herd infection and within-herd prevalence of a disease. In addition, we have presented a plot of the posterior probability of herd infection versus the prior probability of infection, which can be used as both a qualitative and quantitative measure of the likelihood of herd infection based on the observed data and prior input.

APPENDIXES

A. FULL CONDITIONAL DISTRIBUTIONS FOR HYPERGEOMETRIC SAMPLING

Define $\begin{pmatrix} y \\ x \end{pmatrix} = 0$ when $x > y$. We derive the distribution $p(v, x - v, z|\eta, \theta, k)$ as follows:

$$
\begin{aligned}
p(v, x - v, z|\eta, \theta, k) &= p(z|\eta, \theta, k)P(v, x - v|z, \eta, \theta, k) \\
&= p(z|\eta, \theta, k)p(v|z, \eta, \theta, k)p(x - v|z, \eta, \theta, k) \\
&= \frac{\begin{pmatrix} k \\ z \end{pmatrix}\begin{pmatrix} N - k \\ n - z \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}\left\{\begin{pmatrix} z \\ v \end{pmatrix}\eta^v(1 - \eta)^{z-v}\right\} \\
&\quad \times \left\{\begin{pmatrix} n - z \\ x - v \end{pmatrix}(1 - \theta)^{x-v}\theta^{n-z-x+v}\right\}.
\end{aligned}
$$

This leads to the full conditional distributions:

$$
\begin{aligned}
\eta|\text{else} &\sim \text{beta}(a_\eta + v, b_\eta + z - v), \\
\theta|\text{else} &\sim \text{beta}(a_\theta + n - z - x + v, b_\theta + x - v), \\
p(k|\text{else}) &\propto p(k)\begin{pmatrix} k \\ z \end{pmatrix}\begin{pmatrix} N - k \\ n - z \end{pmatrix}, \\
p(z|\text{else}) &\propto \begin{pmatrix} k \\ z \end{pmatrix}\begin{pmatrix} N - k \\ n - z \end{pmatrix}\begin{pmatrix} z \\ v \end{pmatrix}\begin{pmatrix} n - z \\ x - v \end{pmatrix}\left(\frac{1 - \eta}{\theta}\right)^z, \\
p(v|\text{else}) &\propto \begin{pmatrix} z \\ v \end{pmatrix}\begin{pmatrix} n - z \\ x - v \end{pmatrix}\left(\frac{\eta\,\theta}{(1 - \eta)(1 - \theta)}\right)^v.
\end{aligned}
$$

B. FULL CONDITIONAL DISTRIBUTIONS FOR BINOMIAL SAMPLING

The following conditional distributions are straightforward to calculate:

$$
\begin{aligned}
\eta|\text{else} &\sim \text{beta}(a_\eta + v, b_\eta + z - v), \\
\theta|\text{else} &\sim \text{beta}(a_\theta + n - x - z + v, b_\theta + x - v), \\
v|\text{else} &\sim \text{binomial}\left(x, \frac{\eta\pi}{\eta\pi + (1 - \theta)(1 - \pi)}\right), \\
z - v|\text{else} &\sim \text{binomial}\left(n - x, \frac{(1 - \eta)\pi}{(1 - \eta)\pi + \theta(1 - \pi)}\right).
\end{aligned}
$$

To implement a Gibbs sampler we require the full conditional distribution of $\pi|z, \eta, \theta, x$. To this end, note that

$$
\begin{aligned}
P(z = 0|\pi > 0, \eta, \theta) &= \int P(z = 0|\pi, \eta, \theta)P(d\pi|\pi > 0, \eta, \theta) \\
&= \int_0^1 (1 - \pi)^n \left\{ \frac{\Gamma(a_\pi + b_\pi)}{\Gamma(a_\pi)\Gamma(b_\pi)} \pi^{a_\pi - 1}(1 - \pi)^{b_\pi - 1} \right\} d\pi \\
&= \frac{\Gamma(a_\pi + b_\pi)\Gamma(n + b_\pi)}{\Gamma(b_\pi)\Gamma(a_\pi + b_\pi + n)} \overset{\text{def}}{=} g(a_\pi, b_\pi).
\end{aligned}
$$

Because $z = 0$ implies that π is independent of x, using Bayes' rule,

$$
\begin{aligned}
P(\pi = 0|z = 0, \eta, \theta, x) &= P(\pi = 0|z = 0, \eta, \theta) \\
&= \frac{1 - p_I}{1 - p_I + p_I g(a_\pi, b_\pi)} \overset{\text{def}}{=} h(a_\pi, b_\pi, p_I).
\end{aligned}
$$

Finally we have

$$
\begin{aligned}
\pi|z > 0, \eta, \theta, x, z &\sim \text{beta}(a_\pi + z, b_\pi + n - z), \\
\pi|z = 0, \eta, \theta, x &\sim \{1 - h(a_\pi, b_\pi, p_I)\}\text{beta}(a_\pi, b_\pi + n) + h(a_\pi, b_\pi, p_I)\delta_{\{0\}}.
\end{aligned}
$$

ACKNOWLEDGMENTS

[Received November 2001. Revised March 2003.]

REFERENCES

Bulaga, L. L. (1998), "U.S. Voluntary Johne's Disease Herd Status Program for Cattle," in *Proceedings of the Annual Meeting of the U.S. Animal Health Association*, pp. 420–433.

Cameron, A. R., and Baldock, F. C. (1998a), "A New Probability Formula for Surveys to Substantiate Freedom from Disease," *Preventive Veterinary Medicine*, 34, 1–17.

——— (1998b), "Two-Stage Sampling in Surveys to Substantiate Freedom for Disease," *Preventive Veterinary Medicine*, 34, 19–30.

Cannon, R. M. (2001), "Sense and Sensitivity—Designing Surveys Based on an Imperfect Test," *Preventive Veterinary Medicine*, 49, 141–163.

Cannon, R. M., and Roe, R. T. (1982), "Livestock Disease Surveys. A Field Manual for Veterinarians," Bureau of Range Science, Department of Primary Industry, Australian Government Testing Service, Canberra.

Casella, G., and Berger, R. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Duxbury Press.

Collins, M. T., and Sockett, D. C. (1993), "Accuracy and Economics of the USDA-Licensed Enzyme-Linked Immunoassay for Bovine Paratuberculosis," *Journal of the American Veterinary Medical Association*, 203, 1456–1463.

Courtney, C. H., and Zeng, Q. (1993), "Sensitivity and Specificity of Two Heartworm Antigen Detection Tests," *Canine Practitioner*, 18, 20–22.

Dargatz, D. A., Byrum, B. A., Barber, L. K., Sweeney, R. W., Shulaw, W. P., Jacobsen, R. H., and Stabel, J. R. (2001), "Evaluation of a Commercial ELISA for Diagnosis of Paratuberculosis in Cattle," *Journal of the American Veterinary Medical Association*, 218, 1163–1166.

Dubey, J. P., Thulliez, P., Weigel, R. M., Andrews, C. D., Lind, P., and Powell, E. C. (1995), "Sensitivity and Specificity of Various Serologic Tests for Detection of Toxoplasma gondii Infection in Naturally Infected Sows," *American Journal of Veterinary Research*, 56, 1030–1036.

Enøe, C., Georgiadis, M. P., and 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.

Gastwirth, J. L., Johnson, W. O., and Reneau, D. M. (1991), "Bayesian Analysis of Screening Data: Application to AIDS in Blood Donors," *Canadian Journal of Statistics*, 19, 135–150.

Geisser, S., and Johnson, W. O. (1992), "Optimal Administration of Dual Screening Tests for Detecting a Characteristic with Special Reference to Low Prevalence Diseases," *Biometrics*, 48, 839–852.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Greiner, M., and Gardner, I. A. (2000), "Epidemiologic Issues in the Validation of Veterinary Diagnostic Tests," *Preventive Veterinary Medicine*, 45, 3–22.

Hanson, T. E., Johnson, W. O., and Gardner, I. A. (2003), "Hierarchical Models for Estimating Herd Prevalence and Test Accuracy in the Absence of a Gold-Standard," *Journal of Agricultural, Biological, and Environmental Statistics*, 8, 223–239.

Johnson, W. O., and Gastwirth, J. L. (1991), "Bayesian Inference for Medical Screening Tests: Approximations Useful for the Analysis of AIDS Data," *Journal of the Royal Statistical Society*, Series B, 53, 427–439.

Johnson, W. O., Gastwirth, J. L., and Pearson, L. M. (2001), "Screening Without a Gold Standard: The Hui-Walter Paradigm Revisited," *American Journal of Epidemiology*, 153, 921–924.

Joseph, L., Gyorkos, T. W., and Coupal, L. (1995), "Bayesian Estimation of Disease Prevalence and Parameters for Diagnostic Tests in the Absence of a Gold Standard," *American Journal of Epidemiology*, 141, 263–272.

Suess, E. A., Gardner, I. A., and Johnson, W. O. (2002), "Hierarchical Bayesian Model for Prevalence Inferences and Determination of a Country's Status for an Animal Pathogen," *Preventive Veterinary Medicine*, 55, 155–171.

Sweeney, R. W., Whitlock, R. H., and Buckley, C. L. (1995), "Evaluation of a Commercial Enzyme-Linked Immunosorbent Assay for the Diagnosis of Paratuberculosis in Dairy Cattle," *Journal of Veterinary Diagnostic Investigation*, 7, 488–493.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1762.

Whitlock, R. H., Wells, S. J., Sweeney, R. W., and Van Tiem, J. (2000), "ELISA and Fecal Culture for Paratuberculosis (Johne's disease); Sensitivity and Specificity of Each Method," *Veterinary Microbiology*, 77, 387–398.