Biostatistics (2006), **7**, 1, *pp.* 41–57
doi:10.1093/biostatistics/kxi039
Advance Access publication on June 9, 2005

Bayesian inference for prevalence and diagnostic test accuracy based on dual-pooled screening

TIMOTHY E. HANSON*

Division of Biostatistics, Department of Public Health, University of Minnesota, Minneapolis, MN 55455, USA
hanson@biostat.umn.edu

WESLEY O. JOHNSON

Department of Statistics, University of California, Irvine, CA 92697, USA

JOSEPH L. GASTWIRTH

Department of Statistics, George Washington University, Washington, DC 20052, USA

SUMMARY

We propose a useful protocol for the problem of screening populations for low-prevalence characteristics such as HIV or drugs. Current HIV screening of blood that has been donated for transfusion involves the testing of individual blood units with an inexpensive enzyme-linked immunosorbent assay test and follow-up with a more accurate and more expensive western blot test for *only* those units that tested positive. Our cost-effective pooling strategy would enhance current methods by making it possible to accurately estimate the sensitivity and specificity of the initial screening test, and the proportion of defective units that have passed through the system. We also provide a method of estimating the distribution of prevalences for the characteristic throughout the population or subpopulations of interest.

Keywords: AIDS; Bayesian approach; Diagnostic testing; Gibbs sampling; HIV testing; Prevalence; Sensitivity; Specificity.

1. INTRODUCTION

We consider the problem of screening a population for a low-prevalence characteristic. We assume that data are collected at various sites which may have different prevalences. For example, blood donations are collected at various locations throughout the country and it is likely that the prevalence of HIV among blood donors varies by location. We also discuss the situation when the prevalence rates are similar for sites within a region but the regional prevalences vary substantially. Thus, inferences about the corresponding prevalence distributions (PD) are made.

Our focus is on an extension of the 'standard screening protocol', where units are initially screened one at a time. Screened units are indicated to be (+) or (−) and may be incorrectly classified in either case. We assume the existence of a 'gold standard' (GS) test. Positive units are retested with the GS to confirm

*To whom correspondence should be addressed.

their status while negative units are not. Since the negatives are not retested, there is no information for estimating the sensitivity of the test or the prevalence of the characteristic in a screened population (Johnson and Gastwirth, 1991).

To remedy these difficulties, Gastwirth and Johnson (1994) introduced a two-stage protocol whereby a sample of screened negatives from the standard procedure would be grouped and retested and Johnson and Gastwirth (2000) extended this work by allowing for pooling at both stages. Their methods rely on frequentist asymptotic results and only apply to situations where a single homogeneous region has been sampled. Johnson and Pearson (1999) developed a Bayesian approach to the work of Gastwirth and Johnson (1994). Extensive arguments for cost effectiveness of pooling are given in Gastwirth and Johnson (1994) and Johnson and Gastwirth (2000), and their points have been noted by Hughes-Oliver and Rosenberger (2000) and Tebbs and Swallow (2003). Vansteelandt *et al.* (2000) also considered optimal pool size and modeled covariates corresponding to individual pool members, while Xie (2001) developed regression methods for group-tested samples.

Our goal here is to take the two-stage sampling setup introduced by Johnson and Gastwirth (2000) and develop a Bayesian procedure within a hierarchical framework that allows for sampling multiple sites with possibly varying prevalences in a given region and further broaden the model to multiple regions. Formally, we assume multiple-exchangeable populations within a region with nonexchangeability between regions. The method developed does not require large samples for its validity as in Johnson and Gastwirth (2000) and Gastwirth and Johnson (1994), since Bayesian methods do not rely on asymptotics for their validity, and allows for the incorporation of expert knowledge through the specification of the prior distribution. We also avoid the possibly tenuous assumption made in Johnson and Gastwirth (2000) and several of the references above that the sensitivity of a test is the same whether testing individual units or pools. Zenios and Wein (1998) note that the *dilution effect* inherent in group testing for HIV can compromise the sensitivity of a test resulting in the underestimation of prevalence. One realistic simplification that we initially make is that at the second stage we apply a GS test while Johnson and Gastwirth (2000) allowed for an imperfect test at that stage. The availability of a GS test is required for first-stage classification of initially screened positive pools and would certainly be available at the second stage as well. However, an imperfect but inexpensive test may be available at the second stage and this simplification is relaxed in Appendixes D and E, available online at the *Biostatistics* Web site (www.biostatistics.oxfordjournals.org).

Gastwirth (1987), Brookmeyer (1999), and Enøe *et al.* (2000) presented selected surveys of the general screening literature. Related work involving pooling has been done by Gastwirth and Hammick (1989), Chen and Swallow (1990), Hammick and Gastwirth (1994), Hughes-Oliver and Swallow (1994), Litvak *et al.* (1994), Tu *et al.* (1994), Mendoza-Blanco *et al.* (1996), Weinberg and Umbach (1999), and Sham *et al.* (2002). Kendziorski *et al.* (2003) have shown that group testing has applicability in genetic and microarray analyses. Tu *et al.* (1999) developed Markov chain Monte Carlo (MCMC) methods based on Gibbs sampling for general screening problems, including screening for HIV, using pooled single-stage samples. Our work extends theirs to a second stage of pooled sampling for quality control, and also extends asymptotic Bayesian results for single-unit testing (Johnson and Gastwirth, 1991; Gastwirth *et al.*, 1991) to accommodate pooled samples. Most of the previous literature focused on inference for a single population.

The timeliness of the proposed protocol is underscored by a recent outbreak of West Nile virus in the United States, and subsequent pooled screening of donated blood units for this pathogen. Starting in June 2003, blood collection agencies began using an investigational nucleic acid amplification test (NAT) to screen blood donations for the West Nile virus. Approximately 6 million units were screened from June to December 2003, and at least 800 infected units were removed from the blood supply during this time. Samples from either 6 or 16 units were pooled, depending on the kit manufacturer, and tested by NAT. If a pool was positive, individual samples were tested, generally using material sampled from the corresponding unit. Similar to what we propose in this paper, the sensitivity of the pooled NAT screening

protocol was estimated by having selected blood collection agencies perform retrospective testing studies in areas where prevalence was believed to be relatively high. From initially screened negative pools, individual units were selected and retested for the virus (Centers for Disease Control and Prevention, 2004).

In the next section, we present background material and describe our proposed protocol. Methodology is developed in Section 3 and illustrations are presented in Section 4 with final conclusions in Section 5. Some technical details, derivations, and model extensions are given in the appendixes.

2. Development of the statistical model

The data collection procedure and model will first be described for a single region. A total of n_i units are sampled at site i for $i = 1, \ldots, I$. These units are pooled at the site into m_i pools of size k; for convenience, we assume that $m_i = n_i/k$ is an integer. Disease presence (absence) is denoted by D ($\bar{D}$). A screening test is used to determine if each of the m_i pools is 'reactive', which we denote as (+), or not, which is denoted (−). If a pool tests positive, then a GS test will be applied to the pool. If the pool is GS negative, then the units in the pool are considered to be $\bar{D}$, while if the pool is GS positive, individual units in the pool will be tested with the GS to determine the true status of each.

As screened negative pools are not retested, they may not be truly negative. In order to estimate the prevalence among false-negative units and the sensitivity of the screening test, we incorporate a *quality control* step in which units are randomly sampled from the collection of units that were in negative pools. Then these units are subsequently repooled and retested with the GS test. Since the prevalence of disease in the screened negative units should be quite low, the quality control stage pools can be larger than those in the first stage, provided the sensitivity of the GS test is not compromised. As before, individual units in a positive pool will be retested.

The first-stage data observed for the ith site consist of the numbers of true-positive, false-positive, and -negative initial pools, x_i^{tp}, x_i^{fp}, and x_i^{n}, respectively, the number of units in true-positive pools that are actually infected, x_i, and the numbers that are not, $y_i = kx_i^{\mathrm{tp}} - x_i$. The corresponding second-stage data are the number, $r_{il} \geqslant 0$, of units that were resampled from each initial negative pool, indexed by l, and the number, s_{il}, that are actually D. Table 1 illustrates the situation.

While Johnson and Gastwirth (2000) assumed that data would be collected from multiple sites, they treated the data as a random sample from a single population. A more realistic scenario would allow for geographic variation and consider site-specific prevalences, as considered here. Test accuracies (sensitivity and specificity) for the pooled test are assumed to remain constant from site to site. Define these quantities as

$$\eta = P(+|\geqslant 1D), \quad \theta = P(-|\text{no } D\text{s}),$$

and let π_i denote the prevalence at the ith site. The sensitivity can depend on the pool size k and this quantity is defined without reference to the number of Ds there might be in a given pool. Johnson and

Table 1. *Outcomes of the two-stage process for site i*

	D	$\bar{D}$	
First stage			
(+)	x_i^{tp}	x_i^{fp}	
(−)	*	*	x_i^{n}
			m_i
	$x_i^{\mathrm{tp}} \to (x_i, y_i)$		
Second stage	$x_i^{\mathrm{n}} \to (r_{il}, s_{il}),\ l = 1, \ldots, x_i^{\mathrm{n}}$		

Gastwirth (2000) on the other hand assumed the sensitivity to be the same for different pool sizes, say $k = 1, 2, \ldots, 10$. This assumption requires scrutiny, as they point out, but is unnecessary for the model considered here. The test accuracies may well also vary from site to site with the prevalences, but it is expected that this variation will be relatively small for the types of characteristics considered.

The likelihood function for these data is complicated. However, if we had the same information for initially negative pools as we do for the positive ones, this would amount to full information and the likelihood would factor as the equivalent of independent binomial contributions for all parameters. We thus take advantage of this by constructing 'latent' data that result in an augmented data likelihood of this type. Then we implement Bayesian inferences via Gibbs sampling. Details are given in the next section.

We assume that scientific uncertainty about η and θ can be represented by independent beta prior distributions as in Gastwirth *et al.* (1991). Specifically, we model

$$\eta \sim \text{beta}(a_\eta, b_\eta), \quad \theta \sim \text{beta}(a_\theta, b_\theta).$$

Two possibilities for modeling the PD are considered. First, we model knowledge about the prevalences parametrically as

$$\pi_i \overset{\text{iid}}{\sim} \text{beta}(\mu\gamma, (1-\mu)\gamma).$$

We further place a distribution on (μ, γ), where μ is the mean of the PD and γ measures how concentrated the prevalences are about μ since the variance of the PD is $\mu(1-\mu)/(1+\gamma)$.

Our second approach for modeling the PD is nonparametric and thus allows for flexibility beyond that provided by the beta distribution. We use a mixture of Dirichlet processes (MDP) prior (Ferguson, 1973; Antoniak, 1974) for the distribution that governs the π_is. More precisely,

$$\pi_i \overset{\text{iid}}{\sim} G, \quad G|M, \alpha, \beta \sim \text{DP}\{M\,\text{beta}(\mu\gamma, (1-\mu)\gamma)\},$$

with $M \sim \Gamma(a_M, b_M)$. The precision M is often interpreted as a prior sample size given to the parametric family of beta distributions relative to the number of sites I. Escobar and West (1995) discuss the implementation of a gamma prior on M; see Appendix A.

Considering a standard beta(α, β) distribution, Gelman *et al.* (1995, p. 131) developed the relatively noninformative hyperprior for (α, β), $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$. In our parametrization, this prior translates by the usual Jacobian transformation technique to $p(\mu, \gamma) \propto \gamma^{-3/2}$, corresponding to a uniform prior on μ and an independent and improper prior on γ, $p(\gamma) \propto \gamma^{-3/2}$. In the sequel, we allow for the use of an informative beta prior on μ, since scientists will have prior knowledge on the average prevalence for sites in a given region, but use the above improper prior for γ in conjunction with it. When scientific information is also available for a percentile of the PD (in addition to information for μ), a more informative prior (for γ) developed in Hanson *et al.* (2003) can be used.

3. STATISTICAL INFERENCES: GIBBS SAMPLING

In this section, we develop the necessary machinery to implement a Bayesian analysis of data obtained from the sampling scheme and model described in Section 2. We refer to initial pooling of samples as 'first-stage' pooling, and to the subsequent repooling of units selected from first-stage negative pools as 'second-stage' pooling.

Define the latent data z_{il} as the number of Ds that are in the first-stage negative pool l from site i. Let the total number of Ds in the negative groups at site i be denoted by $z_{i\bullet}$, and let $z_{il}^{\text{fn}} = I_{\{>0\}}(z_{il})$ indicate whether or not pool (il) is false negative, and $z_{\bullet\bullet}^{\text{fn}} = \sum_{i=1}^{I} \sum_{l=1}^{x_i^{\text{n}}} z_{il}^{\text{fn}}$, the total number of first-stage false-negative pools. As the prevalence of D in the screened negatives is very low, an infected second-stage pool is almost certain to have only one D, although we do not use this approximation in obtaining inference.

The notation $(\bullet)$ is similarly applied to the formulas for the observed data yielding $x_{\bullet}^{\mathrm{tp}}$, $x_{\bullet}^{\mathrm{fp}}$, and $x_{\bullet}^{\mathrm{n}}$. Then the augmented data likelihood can be shown to be

$$\mathrm{Lik}_{\mathrm{aug}} = \prod_{i=1}^{I} \pi_i^{x_i + z_{i\bullet}} (1-\pi_i)^{n_i - x_i - z_{i\bullet}} \eta^{x_{\bullet}^{\mathrm{tp}}} (1-\eta)^{z_{\bullet\bullet}^{\mathrm{fn}}} \theta^{x_{\bullet}^{\mathrm{n}} - z_{\bullet\bullet}^{\mathrm{fn}}} (1-\theta)^{x_{\bullet}^{\mathrm{fp}}}. \tag{3.1}$$

Denote $\mathbf{z}$ as the latent data, $\mathbf{x}$ the observed data, and $\boldsymbol{\pi}$, the collection $\{\pi_i\}$. Then with our priors specified in the previous section, we have the following independent 'full-conditional' distributions for the sensitivity and specificity

$$\eta|\mathbf{z},\mathbf{x} \sim \mathrm{beta}(a_\eta + x_{\bullet}^{\mathrm{tp}}, b_\eta + z_{\bullet\bullet}^{\mathrm{fn}}), \quad \theta|\mathbf{z},\mathbf{x} \sim \mathrm{beta}(a_\theta + x_{\bullet}^{\mathrm{n}} - z_{\bullet\bullet}^{\mathrm{fn}}, b_\theta + x_{\bullet}^{\mathrm{fp}}). \tag{3.2}$$

We also obtain independent full conditionals for the prevalences under our parametric (independent beta) specification for the π_is, namely

$$\pi_i|\mathbf{z},\mathbf{x},\eta,\theta,\mu,\gamma \overset{\mathrm{ind}}{\sim} \mathrm{beta}(\mu\gamma + x_i + z_{i\bullet}, (1-\mu)\gamma + n_i - x_i - z_{i\bullet}). \tag{3.3}$$

Then we obtain the full conditional for (μ, γ) as

$$p(\mu,\gamma|\boldsymbol{\pi}) \propto \left[\frac{\Gamma(\gamma)}{\Gamma(\mu\gamma)\Gamma(\gamma-\mu\gamma)}\right]^I \prod_{i=1}^{I} \pi_i^{\mu\gamma} (1-\pi_i)^{(1-\mu)\gamma}\, p(\mu,\gamma), \tag{3.4}$$

where $p(\mu, \gamma)$ is one of the forms given in the previous section.

Now consider sampling the latent number of Ds in each initially screened negative pool, z_{il}, which are conditionally independent given (r_{il}, s_{il}), η, θ, and π_i. Let $N_i(r,s)$ denote the number of negative pools at site i having the same $u = (r,s)$ value. Then define $\mathbf{v}_i(r,s) \in \mathbb{R}^{k+1}$ to be the $k+1$ dimensional vector of counts of pools with $(0, 1, \ldots, k)$ Ds, respectively, out of those $N_i(r,s)$ pools. Then

$$\mathbf{v}_i(r,s) \sim \mathrm{multinomial}(N_i(r,s), \mathbf{q}_i(r,s)),$$

where $\mathbf{q}_i(r,s)$ is the vector of probabilities

$$\mathbf{q}_i(r,s) = \begin{bmatrix} P(z_{il}=0|r,s,\boldsymbol{\phi}) \\ P(z_{il}=1|r,s,\boldsymbol{\phi}) \\ \vdots \\ P(z_{il}=k|r,s,\boldsymbol{\phi}) \end{bmatrix},$$

and $\boldsymbol{\phi} = (\eta, \theta, \pi_1, \ldots, \pi_I)'$. For $1 \leqslant j \leqslant k$

$$\begin{aligned} P(z_{il}=j|r,s,\boldsymbol{\phi}) &= \frac{(1-\eta)p(j)q(j;u)I\{s \leqslant j \leqslant k-(r-s)\}}{(1-\eta)\sum_{j=\{1\vee s\}}^{k-(r-s)} p(j)q(j;u) + \theta p(0) I\{s=0\}}, \\ P(z_{il}=0|r,s,\boldsymbol{\phi}) &= \frac{\theta p(0) I\{s=0\}}{(1-\eta)\sum_{j=\{1\vee s\}}^{k-(r-s)} p(j)q(j;u) + \theta p(0) I\{s=0\}}, \end{aligned} \tag{3.5}$$

where

$$p(j) = \binom{k}{j} \pi_i^j (1-\pi_i)^{k-j}, \quad q(j;u) = \binom{j}{s}\binom{k-j}{r-s} \Big/ \binom{k}{r}.$$

The probabilities (3.5) are derived in Appendix B. Note that $z_{i\bullet} = (0, 1, \ldots, k-1, k) \sum_{(r,s)} \mathbf{v}_i(r,s)$, the number of infected, initially screened negative units at site i.

For the model that replaces the beta distribution on the prevalences with a mixture of DP prior, the conditional distributions in (3.3) and (3.4) are replaced by those given in Hanson *et al.* (2003), which are also given in Appendix A.

The distributions for sensitivity, specificity, and the prevalences are readily sampled. The distribution (3.4) for (μ, γ) can be sampled either by Metropolis–Hastings (Tanner, 1996) or by using the full conditionals for $\mu|\gamma$ and for $\gamma|\mu$, obtainable from (3.4), and utilizing the Gilks and Wild (1992) adaptive rejection algorithm.

The Gibbs sampler involves specifying initial values for the sensitivity, specificity, and prevalence parameters. Then the distribution (3.5) is sampled to obtain the current iterate for the latent data. Then the conditionals given in (3.2)–(3.4) are successively sampled until a new iterate has been obtained, and the process is repeated. The resulting iterates are monitored until they stabilize, and after this 'burn-in', the remaining iterates form a Markov chain from the stable distribution, which is the joint posterior for the parameters and latent data. Subsequent inferences are based on these Monte Carlo samples (cf. Gelman *et al.*, 1995; Tanner, 1996). A conservative assessment of convergence is given by inspecting both the history of sampled iterations of each parameter in the model and corresponding 'running quantile plots'. Robert and Casella (1999) discuss convergence diagnostics at length.

The above results are easily extended to a situation with several regions. In this instance, the prevalences from each region have their own distinct PD, which is beta(μ, γ). The parameters (μ, γ) are distinct from region to region. Similarly, in the nonparametric case, the unknown baseline distribution for the prevalences is presumed to be different from one region to the next. We assume independence between regions conditional on the relevant hyperparameters of the model.

As in Johnson and Gastwirth (2000), a basic premise of our work is the existence of a GS test when applied to pooled data at the first stage. The purpose of first-stage pooling is cost effectiveness. In our primary illustration of HIV testing, the western blot is regarded as a GS test. However, the expense associated with using it may only be justifiable in the first stage. This would be especially true if the second stage is performed purely for statistical purposes. Thus, in Appendixes D and E (available at the *Biostatistics* Web site), we explore other possibilities that allow for imperfect testing at the second stage. Two ways of accomplishing this involve taking the second-stage pools already discussed and (1) applying an imperfect diagnostic test to the second-stage pools or (2) testing the individual units corresponding to the second-stage pools with an imperfect diagnostic test. While this second method would not be applied in the case of HIV testing, it would be applicable to situations where it was expected that a moderate to large number of units would escape detection at the first stage. Both generalizations allow for either second-stage accuracies to be known exactly or for the use of prior information. There are numerous other possibilities that we do not consider here but would be straightforward to implement given our development of these two situations. For example, method (1) could be modified to allow for individual imperfect retesting of individual units in second-stage positive pools. We do not pursue this further since we believe that protocol (1) would suffice for statistical purposes based on the work of Johnson and Gastwirth (2000).

4. ILLUSTRATIONS

In this section, we show the procedure works for data that are simulated. Currently, the two-stage pooling procedure is not used in practice so actual data cannot be analyzed. The utility of the method will be demonstrated on simulated data that is constructed with known results from HIV literature. While the illustrations focus on potential HIV studies, the methods apply to any low-prevalence disease/infection or disease gene for which a GS test exists.

We first consider a situation with a single region under a parametric PD assumption and where we vary the negative sampling fraction, f, to see how precision is increased with f. Then we consider a scenario where two 'regions' have been sampled but where the scientists involved have unknowingly treated them as a single region. We use our method for a single population based on having a nonparametric

distribution for the PD and with relatively noninformative priors for all parameters. For such an example, our method should yield a posterior PD reflecting the fact that the data come from a mixture of two distinct subpopulations. This will be seen to be the case.

Our third scenario involves potential HIV testing in the United States where the country is divided into three regions. In this instance, we use the parametric model for the three PDs. Data are simulated using available information about the overall HIV prevalence in blood donors, and using additional information about AIDS prevalence in the three regions and the United States as a whole.

Finally, we perform a modest simulation experiment with $I = 1$ site to demonstrate that the Bayesian estimators provide well-defined credible intervals and coverage rates comparable to or greater than previously published method of moments estimators.

Simulated data generation for all scenarios was random with n units sampled for each of the I sites. For a fixed site with prevalence π, the number of infected units was generated as binomial(n, π). These infected units were then randomly (uniformly) allocated into one of the n/k pools of size $k = 5$, 10, or 20. Each pool was then 'classified' as infected or not with probability η or $1-\theta$ depending on its true infection status. For the positive pools, all units were screened using the GS. For every negative pool, r units were randomly selected and (implicitly pooled into groups) given the GS so $f = r/k$. In the low-prevalence setting if r is 1, then (r, s) is nearly always (1,0), i.e. the probability that a randomly selected unit from a screened negative pool is truly positive is extremely small. In the third example, we set $\eta = 0.98$ and $\theta = 0.99$ as these are close to the actual values for pooled enzyme-linked immunosorbent assay HIV tests. Typically, posterior equal-tailed 95% credible intervals and and posterior medians are reported.

4.1 *Simulated single-region HIV data analyzed as a single region using the parametric model for the PD*

We set $I = 500$ sites with $n_i = 1000$ units sampled at each site. The initial pool size was $k = 20$ and two values of f, 0.05 and 0.5, are considered. The distribution of the π_is was simulated as

$$\pi_1, \ldots, \pi_{500} \overset{\text{iid}}{\sim} \text{beta}(4.405, 98.183).$$

Our priors were $\eta, \theta, \mu \sim \beta(1, 1)$, and $p(\gamma) \propto \gamma^{-3/2}$. Results are given in Table 2. Clearly, the method is able to pin down all the parameters with very high precision. Notice that larger f results in more precise estimation of η while the other parameters are already estimated accurately using the smaller f.

4.2 *Simulated two-region HIV data analyzed as a single region using the mixture of Dirichlet processes model for the PD*

We set $I = 500$ sites with $n_i = 1000$ units at each site. The initial pool size was $k = 10$ and the screened negative repooling fraction was $f = 0.1$. We generated the disease prevalence π_i at site i according to

$$\pi_1, \ldots, \pi_{400} \overset{\text{iid}}{\sim} \text{beta}(0.3, 65), \quad \pi_{401}, \ldots, \pi_{500} \overset{\text{iid}}{\sim} \text{beta}(15, 1000).$$

Roughly, the distribution of the π_is is a mixture of two beta densities.

Table 2. *Posterior estimates for single-region data*

Parameter	Posterior estimates			
	$f = 0.05$		$f = 0.5$	
	Median	95% CI	Median	95% CI
$\eta = 0.98$	0.980	(0.971, 0.987)	0.979	(0.976, 0.982)
$\theta = 0.99$	0.990	(0.988, 0.992)	0.991	(0.989, 0.993)
$\mu = 0.0429$	0.0429	(0.0411, 0.0448)	0.0430	(0.0414, 0.0446)
$\sigma = 0.199$	0.0185	(0.0170, 0.0201)	0.0187	(0.0171, 0.0201)

Table 3. *Posterior test accuracy estimates*

Parameter	Posterior estimates	
	Median	95% CI
$\eta = 0.98$	0.986	(0.963, 0.996)
$\theta = 0.99$	0.990	(0.989, 0.991)

We modeled the PD as a single MDP:

$$\pi_1, \ldots, \pi_{500} | G \overset{\text{iid}}{\sim} G, \quad G|M, \alpha, \beta \sim \text{DP}\{M \,\text{beta}(\mu\gamma, (1-\mu)\gamma)\},$$

and used the uninformative hyperprior

$$p(\mu, \gamma) \propto \gamma^{-3/2}.$$

We modeled the prevalence as if it were one homogeneous population to demonstrate that the MDP model is robust to misspecification of the number of regions.

We placed a gamma prior on the precision with effective support on $(50, 300)$: $M \sim \Gamma(20, 0.1)$. Finally, we used uniform priors on the test accuracies:

$$\eta, \theta \sim \text{beta}(1, 1).$$

The posterior distributions for the test accuracies are both precise and accurate, see Table 3. Figure 1 shows Gaussian kernel-smoothed density estimates of the posterior predictive PD (posterior mean of the PD) denoted as π_{501}, and the actual distribution. The MDP model picks out the mode at $\pi_{501} = 0.014$ admirably, although one can see this predictive density is somewhat 'smoothed' towards the beta base-measure of the DP.

Output for the prevalence among the initially screened negative units at site i was monitored, namely

$$\pi_i^- = P(D|-, \text{site } i) = \frac{\pi_i(1-\eta)}{(1-\eta)(1-(1-\pi_i)^k) + \theta(1-\pi_i)^k}. \tag{4.6}$$

The posterior mean, say $\hat{\pi}_i^-$, is an estimate of π_i^-. In our example, the largest $\hat{\pi}_i^-$ was 0.003, 3 in 1000, and the smallest was 0.0000043, or about 4 in a million. The interval $(4.5/10^6, 5/10^4)$ contains 95% of the $\hat{\pi}_i^-$s.

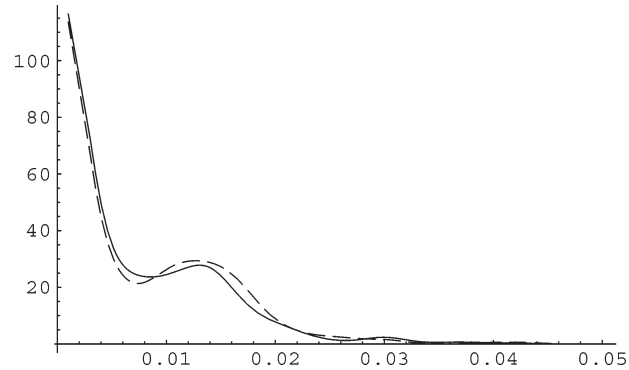

Fig. 1. Posterior predictive prevalence density (solid) and actual density (dashed).

If we had believed *a priori* that this population consisted of several homogeneous subpopulations, we could have modeled each subpopulation separately with a different beta or MDP PD. For example, the United States might be divided into high-, medium-, and low-population density regions in the case of HIV testing for donated blood, as in the next scenario.

4.3 *Testing for HIV in blood donors in three regions of the nation using distinct parametric PDs*

Here, we consider the problem of screening for HIV among blood donors in the United States. It is of interest to assess the first-stage sensitivity of the pooled test as well as its specificity, in addition to saving money by pooling samples. Moreover, we are interested in estimating the PD of HIV infections among blood donors across broad regions of the United States. The Centers for Disease Control actually divides the country into four regions. For the purposes of illustration, we consider three regions accounting for 22 states, (i) Great Plains/Midwest (GPM) states: Idaho, Wyoming, Montana, North Dakota, South Dakota, Nebraska, and Kansas, (ii) west coast and southwest (WSW) states: California, Oregon, Washington, Nevada, Arizona, New Mexico, Colorado, and Utah, and (iii) southern east coastal (SEC) states: Florida, Georgia, South Carolina, North Carolina, Virginia, Maryland, and Delaware.

The overall HIV infection rate in first-time blood donors has been assessed to be 15/100 000 by Glynn *et al.* (2000). Moreover, AIDS rates per 100 000 per year have been assessed to be from 0.5 to 4.3 in the GPM, between 6.5 and 12.5 in the WSW, and between 11.5 and 34.6 (with most over 20) in the SEC states (Centers for Disease Control and Prevention, 2001). These numbers are somewhat proportional to the estimated rates for adults and adolescents living with HIV infection but not AIDS (Centers for Disease Control and Prevention, 2002). Thus, we constructed data from the regions GPM, WSW and SEC to have HIV infection rates that were in accord with these numbers but centered overall at 0.00015. We assumed 20 sites in region GPM, 40 in WSW, and 40 in SEC and prevalences were simulated from beta distributions with parameters (μ, γ) = (22.55, 104232.35), (88.63, 157486.96), and (229.20, 28206.46), respectively, for the three regions, see Figure 2. One thousand pools of size 10 were generated at each site so overall, 1 000 000 'units' were tested. Data were simulated for each region as described in the introduction to this section; we set $f = 0.1$.

We considered an analysis of these data based on an informative prior for the sensitivity and specificity of the pooled test. Specifically, $\eta \sim$ beta(58.4, 1), $\theta \sim$ beta(98.35, 1), $\mu_i \sim$ beta(1, 548.2): $i = 1, 2, 3$. These have modes of 1, 1, and 0, respectively, and 5th percentiles of 0.95 and 0.97, respectively, for η and θ, and 95th percentile of 0.01 for the μ_is. We used our choice of noninformative prior for γ_1, γ_2, and γ_3. The Monte Carlo sample size used was 100 000 and convergence was clearly established.

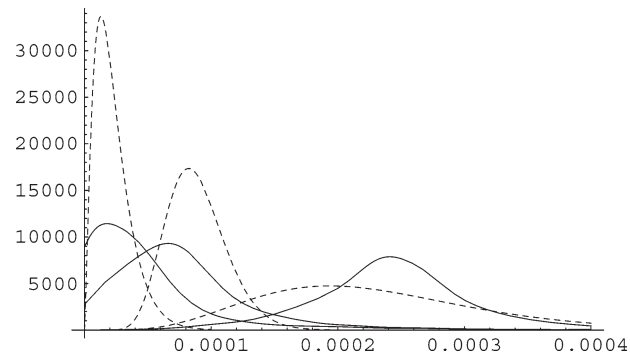

Fig. 2. Posterior predictive prevalence density (solid) and actual densities (dashed).

Estimated versus actual PDs are given in Figure 2. The 'effective' numbers of observations that are available for each region for estimating the PD are only 20, 40, and 40, respectively, so it is not surprising that the estimates are not closer to the true distributions. Moreover, the prevalences are quite small and so with only 10 000 units per site, it would only be possible obtain prevalence estimates on the order of 0.0000, 0.0001, 0.0002, etc. if one knew exactly which units were infected. Nonetheless, the model estimates the distributions reasonably well, given the limited number of sites sampled.

Table 4 gives posterior estimates of the sensitivity and specificity and the average prevalences by region and their differences. The test accuracy estimates are very close to their true values. Since there were only 20 sites in region 1, there is little power to detect the true difference between that region's average prevalence and the others. Credible intervals for γ_1/γ_2, γ_1/γ_3, and γ_2/γ_3 all covered one.

We also tried a slightly different prior on the sensitivity and specificity, $\eta \sim \text{beta}(151.8, 4.08), \theta \sim \text{beta}(212.11, 3.13)$. The mode of the priors for η and θ are now 0.98 and 0.99, respectively, but retain 5th percentiles of 0.95 and 0.97. The resulting inferences for all model parameters are almost identical to those obtained with our original prior; however, the posterior median and credible interval for η were 0.977 (0.947, 0.993), slightly closer to the true value. The posterior estimates for θ were identical under both priors.

The predictive distributions of the numbers of missed units in each region were monitored. The predictive probabilities of missing from 1–3 units out of 200 000, 400 000, and 400 000 for regions 1–3, respectively, are 0.10, 0.27, and 0.47, while the predictive probabilities of catching all the infected units were 0.9, 0.71, and 0.44, respectively. In actuality, 3 units were missed out of 133 infected units at the first stage from the total 1 million units screened across sites.

Potential problems with the estimators proposed in Johnson and Gastwirth (2000) and Gastwirth and Johnson (1994) are illustrated with these data. For this realistic scenario, the prevalences across the $I = 100$ sites are quite low and there is a good chance that a site will have no infected units to begin with. In the data generated for this example, 40 sites have no infected units, and $x_i = 0$ for these sites. Furthermore, $s_{il} = 0$ for all $i = 1, \ldots, 100$ and $l = 1, \ldots, x_i^{\text{n}}$. For the 40 sites with $x_i = 0$, the frequentist estimate of π_i is $\hat{\pi}_i = 0$; see Appendix C. This implies that the estimate for η is $\hat{\eta}_i = 0/0$ and is thus undefined, as are *all three* standard errors $\text{se}(\hat{\eta})$, $\text{se}(\hat{\theta})$, and $\text{se}(\hat{\pi}_i)$. Additionally, for the remaining 60 sites with $x_i > 0$, $\hat{\eta} = 1$, and $\text{se}(\hat{\eta}) = 0$ making any sort of reasonable meta-analysis impossible. That is, using the frequentist approach, 40 sites have undefined estimates of η and an estimate of π_i equaling zero; these same sites also yield undefined standard errors for all three estimators. The remaining sites estimate η to be unity with an estimated variance of zero. We stress that this is not an atypical situation and is similar to the screening of West Nile virus described in the Section 1. The prevalences are too small for the frequentist approach to be useful. While an even greater sampler size is warranted for both approaches, the Bayesian approach lands on its feet because of having reasonable prior information.

Table 4. *Posterior estimates for three-region data*, $(*) \times 10^{-6}$

Parameter	Posterior estimates	
	Median	95% CI
$\eta = 0.98$	0.990	(0.948, 0.9996)
$\theta = 0.99$	0.9901	(0.9895, 0.9907)
$^*\mu_1 = 22.5$	48	(20, 768)
$^*\mu_2 = 88.6$	77	(49, 118)
$^*\mu_3 = 229.2$	246	(198, 306)
$^*(\mu_3 - \mu_1) = 206.7$	195	(−521, 263)
$^*(\mu_3 - \mu_2) = 140.6$	169	(108, 236)
$^*(\mu_2 - \mu_1) = 66.1$	27	(−687, 780)

In other examples, we compared California to New Mexico using a similar simulation based on relative AIDS and HIV rates for those states. We used uniform priors for η and θ and the estimated sensitivity was 0.968 with a credible interval of (0.837, 0.999) based on sampling 10 sites in New Mexico and 40 in California with 1000 pools of size 10 at each site. In this instance, the estimated sensitivity is somewhat shrunk towards the prior mean of 0.5. Consequently, it is important to use a prior that has been carefully selected by the experts. Since the prevalence of HIV in the blood supply is monitored regularly, e.g. monthly and certainly yearly, one can use the posterior distributions for the previous period in order to obtain the prior for analyzing data for the current period. In general, sequential updating of the posterior will mitigate this problem as the original prior will 'wash out' in time. Alternatively, selecting a larger fraction of retested units for the second stage than we did, or increasing the number of units sampled per site, would overcome the problem.

4.4 *Comparison of Bayesian approach with Johnson and Gastwirth (2000) for one site*

To compare the relative performance of the Bayesian approach to the approach proposed in Johnson and Gastwirth (2000) for one site, we performed a small simulation study, found in Table 5. The estimator $\hat{\eta}$ proposed in Johnson and Gastwirth (J–G) can be unity with a corresponding standard error of zero (see Appendix C) with positive probability. This happens frequently in small sample size situations when prevalence is low and the test is highly accurate. In these situations, the Bayesian approach offers valid interval estimates, whereas the frequentist interval is the singleton (1, 1) and cannot 'cover' the true η unless η is unity as well. One could instead consider coverage percentages conditional on $\hat{\eta} < 1$, but this would seem to give an unfair measure of performance. For HIV screening, the prevalence is quite small and test accuracies are close to one. It would often be the case that the approach J–G is unusable unless very large sample sizes are considered.

All simulations use uniform priors on θ and π. The prior for η is either uniform or an informative beta prior with a mean at the true value of η used in simulating data. Overall, both approaches yield unbiased estimates of θ and π, with the mean squared error being about the same from the two approaches. The J–G approach is remarkably unbiased in estimating η, whereas the Bayesian approach using the posterior median typically underestimates η. In low-prevalence situations, there is little information for η and the posterior distribution can be markedly skewed left. We have found that using the posterior mode (found by evaluating the posterior density for η on a grid) instead of the posterior median greatly reduces this bias.

Table 5. *Monte Carlo comparison of Bayesian and asymptotic coverage rates*

n	k	f	π	η	θ	a_η	b_η	Bayesian			J–G		
								π%	η%	θ%	π%	η%	θ%
500	5	0.2	0.10	0.98	0.98	1	1	94	86	95	95	18	70
						98	2	96	100	95			
500	5	0.2	0.10	0.9	0.9	1	1	94	93	95	95	64	93
						18	2	96	100	93			
500	10	1.0	0.05	0.8	0.8	14	6	95	97	95	93	88	92
						16	4	96	99	96			
						18	2	96	90	96			
1000	10	0.1	0.10	0.9	0.9	1	1	93	92	92	94	66	90
10 000	10	0.2	0.01	0.98	0.98	1	1	95	94	94	93	32	94
10 000	20	1.0	0.01	0.98	0.98	1	1	95	96	95	95	83	91
10 000	20	0.05	0.01	0.98	0.98	1	1	90	90	95	93	10	91
						98	2	95	100	94			

The coverage rate for the Bayesian approach is comparable to the J–G approach for π across a variety of scenarios; coverage rates for θ are overall slightly better using the Bayesian model. The Bayesian interval estimates for η are defined and contain η at much higher levels than the J–G approach using the asymptotic variances. To assess the impact of diffuse but accurate sensitivity information on inference, we ran three scenarios twice, one with a uniform prior on η, and another with an informative, but relatively vague (prior weights are worth either 100 or 20 pools) beta prior correctly centered at η. Note that in all three cases, good prior information increases the η coverage rate dramatically to 100% and modestly improves π coverage. One scenario was run with a prior centered at the true value $\eta = 0.8$ and two 'incorrect' priors, centered at 0.7 and 0.9, respectively, all with a prior weight of 20 pools. For the misspecified priors, the Bayesian method still outperforms J–G.

In each simulation, a Monte Carlo sample size of 1000 was generated from the model. Each set of Bayesian estimates were based on an MCMC sample size of 1000 after a burn-in of 100; with one site, model parameters mix very quickly and this was deemed adequate for the purposes of simulation.

A referee has suggested exploring a bootstrap procedure to estimate $\mathrm{se}(\hat{\eta})$ rather than the asymptotic formula in Appendix C. A parametric bootstrap would proceed as follows for the case of a single site with perfect second-stage sensitivity: $\eta_2 = 1$.

In stage 1, there are n sampling units divided into m pools of size k. The observed data for stage 1 are $(x^{\mathrm{tp}}, x^{\mathrm{fp}}, x^{\mathrm{n}})$, the number of groups out of m that are true-positive, false-positive, and -negative, respectively. We see x Ds and y $\bar{D}$s out of the kx^{tp} individuals in the true-positive pools. We also see $s_\bullet$ Ds total out of the kx^{n} individuals in the initially screened negative pools, using the GS test at the second stage. Individuals are randomly sampled from initially screened negative pools with fraction f. The J–G estimates of the first- and second-stage prevalences and the sensitivity of the first-stage test are thus

$$\hat{\pi} = \frac{x + s_\bullet/f}{n}, \quad \hat{\pi}_1 = \frac{s_\bullet}{fkx^{\mathrm{n}}}, \quad \hat{\eta} = \frac{x}{n\hat{\pi}}.$$

The bootstrap procedure iterates the following sequence of steps:

1. Sample $(x^{\mathrm{tp}*}, x^{\mathrm{fp}*}, x^{\mathrm{n}*}) \sim \mathrm{multinomial}(m, (x^{\mathrm{tp}}, x^{\mathrm{fp}}, x^{\mathrm{n}})/m)$.
2. Sample $x^* \sim \mathrm{Bin}(kx^{\mathrm{tp}}, x/(x+y))$.
3. Sample $x_s^* \sim \mathrm{Bin}(kx^{\mathrm{n}}, f)$ and $s^* \sim \mathrm{Bin}(x_s^*, \hat{\pi}_1)$.
4. Compute $\hat{\eta}^* = x^*/(x^* + s^*/f)$.

Note that this can fail in two ways: if $s_\bullet = 0$, then $\hat{\pi}_1 = 0$ so $s^* = 0$ and $\hat{\eta}^* = 1$. This can happen with a highly sensitive but imperfect first-stage test. The second situation is when $x = 0$, which happens only if $x^{\mathrm{tp}} = 0$, which would happen in very low-prevalence situations. So the bootstrap procedure fails exactly when the original standard errors are undefined.

A referee has suggested the possibility of extending the original J–G approach to estimate prevalences in multiple regions. Since the problem of undefined estimates using the J–G approach is likely to occur in many instances, the Bayesian approach seems simpler to implement, and has the advantage of allowing for the incorporation of scientific information and for easy sequential updating of the posterior through time.

For a single site, *Mathematica* (Wolfram Research, Inc.) code takes on the order of 30 s for 10 000 pools of size 20 on a 2.8 GHz Pentium 4 with 512 MB of RAM. This involves taking 10 000 MCMC iterates, considerably more than what is needed, after a burn-in of 1000. Programing the model in a compiled language such as FORTRAN is sure to reduce the computing time to less than a second. Fitting multiple sites requires substantially more computation, mainly because the μ and γ MCMC iterates can be highly autocorrelated, although computing times in *Mathematica* are generally on the order of an hour or less. This time would be reduced to minutes in a compiled language. Given that the data take considerable time and effort to collect, the additional time required to fit the Bayesian model should not pose a major inconvenience.

5. CONCLUSIONS

We presented cost-effective means for obtaining information about the sensitivity of a pooled screening test and the prevalence of the characteristic among screened negatives. This is particularly important since the quality of the screening test used in the field ultimately determines the quality of the blood supply. It is well known that the accuracy of tests in practice is often less than in the laboratory (Gastwirth, 1987). Our protocol makes it possible to detect a decline in the accuracy of the screening test using the predictive distribution as in Gastwirth *et al.* (1991). In this context, it is also very important to accurately quantify the prevalence of HIV-infected blood units available for transfusion. The method extends the use of group testing to detect heterogeneous subpopulations and offers the ability to model prevalences as exchangeable and heterogeneous within known subpopulations. The nonparametric version of the method is particularly attractive since it allows for a possibly multimodal PD as can be seen in our second example.

The Bayesian model avoids problems with undefined parameter estimates associated standard errors encountered with the estimators proposed in Gastwirth and Johnson (1994) and Johnson and Gastwirth (2000). These problems can occur in moderate to large sample sizes in rather typical settings involving low prevalence and very accurate tests. The advantages of using good prior information, when available, are evident from the examples.

A number of issues deserve further study. The pool size may affect the sensitivity and the specificity of a screening test. It will be important to determine an optimal pool size for which the first-stage procedure remains highly effective in screening the population and appreciably less expensive than individual testing. While retesting the pools of first-stage negative units at their original collection site may be convenient, it may be preferable to retest them at a central processing location to monitor the testing process at individual sites.

It is also worth considering the feasibility of retesting all of the first-stage negatives in 'large' pools in order to make up for any loss in sensitivity due to pooling at the first stage. Determining the size of the second-stage pool so that the sensitivity of the GS test for the pool remains at or near one will also require study. Other scenarios beyond what is presented in Appendixes D and E for using an imperfect but less expensive second-stage test, considered by Johnson and Gastwirth (2000), also warrants further exploration.

APPENDIX A: MDP FULL-CONDITIONAL DISTRIBUTIONS

We define $\boldsymbol{\pi}^{(i)}$ to be the vector $\boldsymbol{\pi}$ with the ith component removed. Using Bayes' rule, we compute

$$[\pi_i|\boldsymbol{\pi}^{(i)}, \mathbf{z}, \mathbf{x}, \alpha, \beta] \begin{Bmatrix} \sim \text{beta}(x_i + z_{i\bullet} + \alpha, n_i - x_i - z_{i\bullet} + \beta), \text{ with probability } q_i \\ = \pi_j,\ j \neq i, \text{ with probability } q_j \end{Bmatrix},$$

where

$$\hat{q}_i = \frac{M\Gamma(\alpha+\beta)\Gamma(x_i + z_{i\bullet} + \alpha)\Gamma(n_i - x_i - z_{i\bullet} + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n_i + \alpha + \beta)},$$

$$\hat{q}_j = \pi_j^{x_i + z_{i\bullet}}(1 - \pi_j)^{n_i - x_i - z_{i\bullet}}, \quad \text{for } j \neq i\ ,$$

and $q_k = \hat{q}_k / \sum_{s=1}^{I} \hat{q}_s$. Let $M \sim \Gamma(a_M, b_M)$ and let $d(\boldsymbol{\pi})$ denote the number of distinct values in the vector $\boldsymbol{\pi}$. Using a technique from Escobar and West (1995), we introduce the auxillary variable σ and have

$$[\sigma|M] \sim \text{beta}(M + 1, I),$$

$$[M|\sigma, \pi] \sim p\Gamma(a_M + d(\boldsymbol{\pi}), b_M - \log\sigma) + (1 - p)\Gamma(a_M + d(\boldsymbol{\pi}) - 1, b_M - \log\sigma),$$

where

$$p = p(a_M, b_M, \sigma, \pi) = \left(\frac{a_M + d(\boldsymbol{\pi}) - 1}{K(b_M - \log \sigma) + a_M + d(\boldsymbol{\pi}) - 1} \right).$$

Alternatively, M can be left fixed at a prespecified value.

Antoniak (1974) argues

$$[\alpha, \beta | \boldsymbol{\pi}] \sim f(\alpha, \beta | \boldsymbol{\pi}) \propto f(\alpha, \beta) \prod_{i=1}^{d(\pi)} g_{\alpha,\beta}(\pi_i^*),$$

where $\{\pi_i^*\}_{i=1}^{d(\boldsymbol{\pi})}$ is the set of distinct values of $\boldsymbol{\pi}$ and $g_{\alpha,\beta}(\cdot)$ is a beta(α, β) density. Other parameters are sampled as they were in the parametric case.

APPENDIX B: DERIVATION OF (5)

For notational simplicity, we drop the subscripts denoting site i and initially screened negative pool l. Recall that $(-)$ denotes the event that a pool initially screens negative. Define $\binom{n}{m} = 0$ when $m > n$. Let $\boldsymbol{\phi} = (\eta, \theta, \pi_1, \ldots, \pi_I)'$. Then

$$\begin{aligned} p(z|-, \phi) &= \frac{P(-|z, \boldsymbol{\phi}) p(z|\phi)}{P(-|\boldsymbol{\phi})} \\ &= \frac{[\theta I\{z = 0\} + (1 - \eta) I\{z \geqslant 1\}] \binom{k}{z} \pi (1 - \pi)^{k-z}}{P(-|\boldsymbol{\phi})}. \end{aligned}$$

Also,

$$p(z|s, -, \boldsymbol{\phi}) = \frac{p(s|z, -, \boldsymbol{\phi}) p(z|-, \boldsymbol{\phi})}{p(s|-, \boldsymbol{\phi})} = \frac{\binom{z}{s} \binom{k-z}{r-s} p(z|-, \boldsymbol{\phi})}{\binom{k}{r} p(s|-, \boldsymbol{\phi})}.$$

These results yield (3.5), as both $P(-|\boldsymbol{\phi})$ and $p(s|-, \boldsymbol{\phi})$ are free of z.

APPENDIX C: ESTIMATORS AND ASYMPTOTIC VARIANCES FOR FRACTION f OF INITIALLY SCREENED NEGATIVE UNITS RETESTED WITH GS TEST

As only one site is considered, the subscript i is dropped. There are n units divided into m first-stage pools of size k. There are x infected units in the true-positive pools at the first stage. s_l out of r_l are D in each pool $l = 1, \ldots, x^{\mathrm{n}}$; $s_\bullet = \sum_{l=1}^{x^{\mathrm{n}}} s_l$. Analogous estimators to those found in Johnson and Gastwirth (2000) are

$$\hat{\pi} = \frac{x + s_\bullet / f}{n}, \quad \hat{\eta} = \frac{x}{n\hat{\pi}}, \quad \hat{\theta} = 1 - \frac{x^{\mathrm{fp}}}{m(1 - \hat{\pi})^k}.$$

The asymptotic variances are estimated by the following expressions divided by the number of pools m:

$$\hat{\sigma}_\pi^2 = \frac{\hat{\pi}\,(1-\hat{\pi}\,+(1-\hat{\eta})(1/f\,-1))}{k},$$

$$\hat{\sigma}_\eta^2 = \frac{\hat{\eta}(1-\hat{\eta})(1+\hat{\eta}(1/f\,-1)+\hat{\pi}\,(k-1))}{k\hat{\pi}},$$

$$\hat{\sigma}_\theta^2 = \frac{\hat{\theta}(1-\hat{\theta})}{1-\hat{\pi}} + \frac{k(1-\hat{\theta})^2\hat{\pi}\,(1-\hat{\eta})(1/f\,-1)}{(1-\hat{\pi}\,)^2}$$

$$+ \frac{1-\hat{\theta}}{1-\hat{\pi}}\left\{\frac{1}{(1-\hat{\pi}\,)^{k-1}} - 1 - (1-\hat{\theta})\hat{\pi}\,(k-1)\right\}.$$

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.

Brookmeyer, R. (1999). Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics* **55**, 608–612.

Centers for Disease Control and Prevention (2001). *HIV/AIDS Surveillance Report, 2001*, Volume 13. Atlanta: CDC.

Centers for Disease Control and Prevention (2002). *HIV/AIDS Surveillance Report, 2002*, Volume 14. Atlanta: CDC.

Centers for Disease Control and Prevention (2004). Update: West Nile virus screening of blood donations and transfusion-associated transmission—United States, 2003. *Morbidity and Mortality Weekly Report 2004*, Volume 53. Atlanta: CDC, pp. 281–284.

Chen, L. C. and Swallow, W. H. (1990). Using group testing to estimate a proportion, and to test the binomial model. *Biometrics* **46**, 1035–1046.

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.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Gastwirth, J. L. (1987). The statistical precision of medical testing procedures: applications to polygraph/AIDS antibodies test data. *Statistical Science* **2**, 213–238.

Gastwirth, J. L. and Hammick, P. A. (1989). Estimation of the prevalence of rare disease preserving the anonymity of the subjects by group testing: application to estimating the prevalence of AIDS antibodies in blood donors. *Journal of Statistical Planning and Inference* **22**, 15–27.

Gastwirth, J. L. and Johnson, W. O. (1994). Screening with cost-effective quality control: potential applications to HIV and drug testing. *Journal of the American Statistical Association* **89**, 972–981.

Gastwirth, J. L., Johnson, W. O. and Reneau, D. M. (1991). Bayesian analysis of screening data: applications to AIDS in blood donors. *Canadian Journal of Statistics* **19**, 135–150.

GELMAN, A., CARLIN, J. B., STERN, H. S. AND RUBIN, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

GILKS, W. R. AND WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

GLYNN, S. A., KLEINMAN, S. H., SCHREIBER, G. B., BUSCH, M. P., WRIGHT, D. J., SMITH, J. W., NASS, C. C. AND WILLIAMS, A. E. (2000). Trends in incidence and prevalence of major transfusion-transmissible viral infections in US blood donors, 1991–1996. *Journal of the American Medical Association* **294**, 229–235.

HAMMICK, P. A. AND GASTWIRTH, J. L. (1994). Extending the applicability of estimation of the prevalence of sensitive characteristics by group testing to moderate prevalence populations. *International Statistical Review* **62**, 319–331.

HANSON, T., 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**, 223–229.

HUGHES-OLIVER, J. M. AND ROSENBERGER, W. F. (2000). Efficient estimation of the prevalence of multiple rare traits. *Biometrika* **87**, 315–327.

HUGHES-OLIVER, J. M. AND SWALLOW, W. H. (1994). A two-stage adaptive group-testing procedure for estimating small proportions. *Journal of the American Statistical Association* **89**, 982–993.

JOHNSON, W. O. AND GASTWIRTH, J. L. (1991). Bayesian inference for medical screening tests: approximations useful for the analysis of Acquired Immune Deficiency Syndrome. *Journal of the Royal Statistical Society Series B* **53**, 427–439.

JOHNSON, W. O. AND GASTWIRTH, J. L. (2000). Dual group screening. *Journal of Statistical Planning and Inference* **83**, 449–473.

JOHNSON, W. O. AND PEARSON, L. M. (1999). Dual screening. *Biometrics* **55**, 867–873.

KENDZIORSKI, C. M., ZHANG, Y., LAN, H. AND ATTIE, A. D. (2003). The efficiency of pooling mRNA in microarray experiments. *Biostatistics* **4**, 465–477.

LITVAK, E., TU, X. M. AND PAGANO, M. (1994). Screening for the presence of disease by pooling sera samples. *Journal of the American Statistical Association* **89**, 424–434.

MENDOZA-BLANCO, J. R., TU, X. M. AND IYENGAR, S. (1996). Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: applications to HIV screening. *Statistics in Medicine* **15**, 2161–2176.

ROBERT, C. P. AND CASELLA, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

SHAM, P., BADER, J. S., CRAIG, I., O'DONOVAN, M. AND OWEN, M. (2002). DNA pooling: a tool for large-scale association studies. *Nature Reviews Genetics* **3**, 862–871.

TANNER, M. A. (1996). *Tools for Statistical Inference*, 3rd edition. New York: Springer.

TEBBS, J. M. AND SWALLOW, W. H. (2003). Estimating ordered binomial proportions with the use of group testing. *Biometrika* **90**, 471–477.

TU, X. M., KOWALSKI, J. AND JIA, J. (1999). Bayesian analysis of prevalence with covariates using simulation-based techniques: applications to HIV screening. *Statistics in Medicine* **18**, 3059–3073.

TU, X. M., LITVAK, E. AND PAGANO, M. (1994). Studies of AIDS and HIV surveillance screening tests: can we get more by doing less? *Statistics in Medicine* **13**, 1905–1919.

VANSTEELANDT, S., GOETGHEBEUR, E. AND VERSTRAETEN, T. (2000). Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics* **56**, 1126–1133.

WEINBERG, C. R. AND UMBACH, D. M. (1999). Using pooled exposure assessment to improve efficiency in case–control studies. *Biometrics* **55**, 718–726.

XIE, M. G. (2001). Regression analysis of group testing samples. *Statistics in Medicine* **20**, 1957–1969.

ZENIOS, S. A. AND WEIN, L. M. (1998). Pooled testing for HIV prevalence estimation: exploiting the dilution effect. *Statistics in Medicine* **17**, 1447–1467.

[*Received January 9, 2004; first revision January 6, 2005; second revision May 27, 2005; accepted for publication June 1, 2005*]