Published in final edited form as:

Stat Med. 2018 March 30; 37(7): 1101–1114. doi:10.1002/sim.7568.

Prevalence estimation when disease status is verified only among test positives: applications in HIV screening programs

E.G. Thomas^{a,*}, S. Peskoe^a, and D. Spiegelman^{a,b}

^aDepartment of Biostatistics, Harvard School of Public Health, Harvard University, Boston, MA 02115

^bDepartments of Epidemiology and Nutrition and Global Health, Harvard School of Public Health, Harvard University, Boston, MA 02115

Abstract

The first goal of the United Nations' 90–90–90 HIV/AIDS elimination strategy is to ensure that, by 2020, 90% of HIV positive people know their HIV status. Estimating the prevalence of HIV among people eligible for screening allows assessment of the number of additional cases that might be diagnosed through continued screening efforts in this group. Here, we present methods for estimating prevalence when HIV status is verified by a gold standard only among those who test positive on an initial, imperfect screening test with known sensitivity and specificity. We develop maximum likelihood estimators and asymptotic confidence intervals for use in two scenarios: when the total number of test negatives is known (Scenario 1) and unknown (Scenario 2). We derive Bayesian prevalence estimators to account for non-negligible uncertainty in previous estimates of the sensitivity and specificity. The Scenario 1 estimator consistently outperformed the Scenario 2 estimator in simulations, demonstrating the utility of recording the number of test negatives in public health screening programs. For less accurate tests (sensitivity and specificity <90%), the performance of the two estimators was comparable, suggesting that, under these circumstances, prevalence can still be estimated with adequate precision when the number of test negatives is unknown. However, use of the Bayesian approach to account for uncertainty in the sensitivity and specificity is especially recommended for the Scenario 2 estimator, which was particularly sensitive to misspecification of these values. R code for implementing these methods is available at hsph.harvard.edu/donna-spiegelman/software.

Keywords

prevalence estimation; screening test; diagnostic test; partial verification; HIV

1. Introduction

In 2014, the United Nations Programme on HIV/AIDS (UNAIDS) set a new target in the battle to end the global AIDS epidemic. The 90–90–90 strategy aims to ensure that, by 2020, (1) 90% of individuals living with HIV will know their HIV status; (2) 90% of those

*Correspondence to: Emma Thomas, Department of Biostatistics, Harvard School of Public Health, Harvard University, Boston, MA 02115. emmathomas@g.harvard.edu.

diagnosed will be on sustained antiretroviral therapy (ART), and; (3) 90% of those enrolled in ART will be virally suppressed [1]. Modelling suggests that achieving these ambitious goals will be sufficient to end the HIV/AIDS epidemic [2]. In pursuit of the first of these aims, significant resources are required to scale up HIV screening programs. The prevalence of HIV among people eligible for screening is an important measure of the potential value of a screening program, since it enables assessment of the number of additional HIV positives that could be diagnosed through continued screening efforts in this group.

Because of the expense of testing (or screening; we use these terms interchangeably throughout this paper) based on viral load measurement or Western blot (the gold standard methods), HIV screening often consists of one or more cheap and fast tests with imperfect accuracy, known as rapid tests [3]. In some programs, true HIV status for those who test positive is then verified via a gold standard. Since verification is only performed for test positives, data of this type are comparatively easy and cost effective to collect.

In this paper, we address prevalence estimation using data from a single screening test with gold standard verification among test positives, when the sensitivity and specificity of the test are known or have previously been estimated. The latter assumption is appropriate in the HIV context since, due to the high stakes of HIV testing, the tests used in the field are usually well-validated, so that good estimates of the sensitivity and specificity from previous studies are available. We consider two scenarios: (1) when the number of test negatives is known, and (2) when the number of test negatives is unknown. Scenario 2 occurs frequently in practice, and in fact is the norm throughout a large number of programs in Sub-Saharan Africa supported by the United States President's Emergency Plan for AIDS Relief (PEPFAR). The focus in these resource-limited settings is nearly exclusively on screening to identify HIV positive people and enrol them into care, and records are typically not kept on patients who test negative. Interest in estimating prevalence may then arise retrospectively, and is a secondary although increasingly important concern for these programs in light of the global effort to end the AIDS epidemic by 2030 [4].

Previous work on prevalence estimation with gold standard verification among test positives has focused on cases in which the sensitivity and specificity of the screening test(s) are unknown and must be estimated (e.g., [5,6,7, 8,9,10,11]). These methods are applicable when two or more initial screening tests are performed, followed by gold standard verification among those who test positive on at least one of the tests. In this situation, the parameters are not identifiable without additional assumptions about the dependency structure of the screening test results [11]. In the case of a single screening test, a common approach is to use the proportion testing positive to estimate prevalence. However, this estimator is biased, with the severity of the bias depending on the true prevalence and quality of the screening test. If the sensitivity and specificity of the test are known, methods are available to correct for this bias [12], although to our knowledge there are no equivalent published methods that also incorporate information on disease verification in test positives.

The aims of this paper are to develop methods for estimating disease prevalence using screening data from a single test with gold standard verification in test positives only, and when previous estimates of the sensitivity and specificity of the test are available. R code for

applying these methods is provided at hsph.harvard.edu/donna-spiegelman/software. This paper is divided into four sections. We first introduce statistical models for the data generating mechanism and show how to estimate the prevalence via frequentist and Bayesian methods when the number of test negatives is (1) known or (2) unknown. Second, we investigate the performance of the frequentist estimators. Third, we provide data examples. Finally, we describe appropriate contexts for use of these estimators and discuss their strengths and limitations.

2. Statistical models and estimation

Throughout this paper, we use lower case letters to denote fixed parameters, and upper case letters to denote random variables. Let p be the prevalence of HIV in the target population; here, the population of people targeted for screening or for recruitment into a particular study. Table 1 shows the complete data when the results of both a screening test Z_i (0 if individual i screens negative, 1 if i screens positive) and gold standard verification X_i (0 if i is HIV negative, 1 if HIV positive) are known for all individuals screened. We assume that the (Z_i, X_i) are independent and identically distributed for all i . N_{xz} denotes the number of individuals with results $X_i = x$ and $Z_i = z$, $N_{x.} = \sum_z N_{xz}$ and $N_{.z} = \sum_x N_{xz}$. Let the sensitivity of the test be $\theta = \mathbb{P}(Z_i = 1 | X_i = 1)$, denoting the probability of testing positive for true positives, and let the specificity be $\varphi = \mathbb{P}(Z_i = 0 | X_i = 0)$, denoting the probability of testing negative for true negatives.

When θ and φ are known, Rogan and Gladen [12] showed that an unbiased estimator of p based only on the number of test negatives and test positives (i.e., when no gold standard verification has been performed) is

$$\hat{p}_R = \frac{\frac{N_{.1}}{n} + \varphi - 1}{\theta + \varphi - 1}.$$

The variance of $\hat{p}_R$ is σ_R^2/n , where

$$\sigma_R^2 = \text{Var}\left[\frac{N_{.1}}{\theta + \varphi - 1}\right] = \frac{(\theta p + (1 - \varphi)(1 - p))((1 - \theta)p + \varphi(1 - p))}{(\theta + \varphi - 1)^2}$$

They also provide an estimator for p that incorporates uncertainty in the available estimates of θ and φ , $\hat{\theta}$ and $\hat{\varphi}$. This is achieved by replacing θ and φ in Equation 1 with their respective estimators, and estimating the variance of $\hat{p}_R$ accounting for uncertainty in $\hat{\theta}$ and $\hat{\varphi}$ via the delta method.

Here, we address the situation in which gold standard verification is performed among test positives. We present methods for estimating p under two scenarios: (1) when the total number of test negatives, $N_{0.}$, is known (Section 2.1.1), and; (2) when $N_{0.}$ is unknown (Section 2.1.2). In Section 2.1, we discuss frequentist estimation of p when θ and φ are known with negligible error, so that p is the only unknown parameter. In Section 2.2, we

discuss Bayesian estimation of p when θ and φ are not known precisely but the results of previous studies can be used to construct informative prior distributions.

2.1. Frequentist approach

2.1.1. Scenario 1: Number of test negatives is known—Suppose that we observe the data $(Z_i = z_i, X_i Z_i = c_i)$ for all n individuals in the study, so that gold standard HIV status X_i is unknown for individuals who test negative ($Z_i = 0$). That is, we observe $N_{\cdot 0}$ (the number of test negatives), N_{01} (the number of false positives) and N_{11} (the number of true positives). The likelihood is

$$\begin{aligned}\mathfrak{L}_1(p) &= \prod_{i=1}^n \mathbb{P}(Z_i = z_i, X_i Z_i = c_i) \\ &= \prod_{i=1}^n \mathbb{P}(Z_i = 0)^{1-z_i} [\mathbb{P}(X_i = 1, Z_i = 1)^{c_i} \mathbb{P}(X_i = 1, Z_i = 1)^{1-c_i}]^{z_i} \\ &= \prod_{i=1}^n [(1-\theta)p + \varphi(1-p)]^{1-z_i} \left[(\theta p)^{c_i} ((1-\theta)(1-p))^{1-c_i} \right]^{z_i} \\ &= [(1-\theta)p + \varphi(1-p)]^{N_{\cdot 0}} (\theta p)^{N_{11}} ((1-\theta)(1-p))^{N_{01}}.\end{aligned}$$

It is straightforward to maximise the above likelihood and obtain the maximum likelihood estimate (MLE) of p and an asymptotic confidence interval (CI). The MLE and its asymptotic variance are given here and derivations are in the online supporting material, Section 1.1. The MLE is

$$\hat{p}_1 = \frac{-B - \sqrt{B^2 - 4AC}}{2A},$$

where

$$\begin{aligned}A &= -(1-\theta-\varphi)n \\ B &= (1-\theta-\varphi)(N_{\cdot 0} + N_{11}) - \varphi(N_{11} + N_{01}) \\ C &= N_{11}\varphi.\end{aligned}$$

An asymptotic $(1 - \alpha)$ 100% CI based on the Fisher information is

$$\hat{p}_1 \pm \frac{z_{1-\alpha/2} \hat{\sigma}_1}{\sqrt{n}}$$

where

$$\hat{\sigma}_1^2 = \left[\frac{\theta}{\hat{p}_1} + \frac{1-\varphi}{1-\hat{p}_1} + \frac{(1-\theta-\varphi)^2}{(1-\theta)\hat{p}_1 + \varphi(1-\hat{p}_1)} \right]^{-1}$$

and $z_{1-\alpha/2}$ is the $(1-\alpha/2)100^{th}$ percentile of the standard normal distribution.

2.1.2. Scenario 2: Number of test negatives is unknown—Under Scenario 2, we recruit a fixed number $m = N_{\cdot 1}$ of test positives and verify their HIV statuses via a gold standard, so that the data are $(X_i = x_i | Z_i = 1)$ for $i = 1, \dots, m$. That is, we observe N_{01} and N_{11} , with $m = N_{01} + N_{11}$. Note that the fixed sample size m refers to the number of test positives only, and thus differs from the fixed sample size n in Scenario 1, which refers to the total number tested. In this scenario, no information on test negatives is observed. The likelihood becomes

$$\begin{aligned} \mathfrak{L}_2(p) &= \prod_{i=1}^m \mathbb{P}(X_i = x_i | Z_i = 1) \\ &= \prod_{i=1}^m \mathbb{P}(X_i = 1 | Z_i = 1)^{x_i} \mathbb{P}(X_i = 0 | Z_i = 1)^{1-x_i} \\ &= \left[\frac{\theta p}{\theta p + (1-\varphi)(1-p)} \right]^{N_{11}} \left[1 - \frac{\theta p}{\theta p + (1-\varphi)(1-p)} \right]^{m-N_{11}}. \end{aligned}$$

The MLE for p is then:

$$\hat{p}_2 = \frac{(1-\varphi)N_{11}}{(1-\varphi)N_{11} + \theta N_{01}}.$$

An asymptotic $(1-\alpha)$ 100% CI based on the Fisher information is

$$\hat{p}_2 \pm \frac{z_{1-\alpha/2} \hat{\sigma}_2}{\sqrt{m}}$$

where

$$\hat{\sigma}_2^2 = [\theta \hat{p}_2 + (1-\varphi)(1-\hat{p}_2)] \left[\frac{\theta}{\hat{p}_2} + \frac{1-\varphi}{1-\hat{p}_2} - \frac{(\theta+\varphi-1)}{\theta \hat{p}_2 + (1-\varphi)(1-\hat{p}_2)} \right]^{-1}.$$

Derivations for $\hat{p}_2$ and its variance σ_2^2 are given in the online supporting material, Section 1.2.

Note that $\hat{\sigma}_2^2 = 0$ if $\hat{p}_2 = 0$ or 1. By re-arranging Equation 7, it can be shown that $\hat{p}_2 = 1$ when $N_{01} = 0$, and $\hat{p}_2 = 0$ when $N_{11} = 0$. Therefore, if there are no false positives (or true positives — although the latter is extremely unlikely in practice) in the data, the Scenario 2 estimator

will always be equal to zero (or one) with a CI of length zero, and is thus not a useful estimator. This is not surprising because the underlying idea here is to use only the false positive rate (N01 /m) to estimate the prevalence, so that when no false positives are observed there is not enough information to estimate prevalence.

2.2. Bayesian approach

Suppose now that θ and φ are not known precisely, but that previous studies have produced estimates with some degree of uncertainty. Let Θ , Φ and P be random variables representing the unknown θ , φ and p respectively, and let D_s represent the data available under Scenario s ($s = 1, 2$). The likelihood under Scenario s , $(L_s(P, \theta, \varphi | D_s) dP d\theta d\varphi)$, is given by Equations 2 and 6, after replacing the fixed parameter values p , θ and φ with their random counterparts P , Θ and Φ . For a given choice of the prior distribution $\pi(P, \Theta, \Phi)$, the kernel of the posterior distribution will be

$$\pi_s(P, \theta, \varphi | D_s) \propto \pi(P, \theta, \varphi) L_s(P, \theta, \varphi | D_s).$$

The normalizing constant $c = \int \pi(P, \theta, \varphi) L_s(P, \theta, \varphi | D_s) dP d\theta d\varphi$ is typically not required in order to sample from the posterior using Markov Chain Monte Carlo (MCMC) techniques, and so we do not show it here.

Since the parameter vector (p, θ, φ) is not identifiable in a frequentist setting, informative priors for Θ and Φ are required to permit Bayesian estimation of P . Below, we discuss two options for $\pi(P, \Theta, \Phi)$, both of which lead to intractable posterior distributions under both scenarios ($s = 1, 2$). We therefore use MCMC to obtain samples from the posterior distribution. Examples illustrating this procedure are given in Section 4.

2.2.1. Possible prior distributions—Suppose prior information on Θ and Φ is available from an independent study in the form of frequentist point estimates, $\hat{\theta}$ and $\hat{\varphi}$, and their standard errors, $\hat{\sigma}_{\theta}$ and $\hat{\sigma}_{\varphi}$. Our aim is to use this information to specify priors for Θ and Φ . Two options for achieving this are discussed below. We assume that the priors for Θ and Φ are independent, since these proportions would have been estimated in independent samples (a sample of test positives and a sample of test negatives, respectively). We select a uniform prior for P , as this is the main parameter of interest and we hence prefer that its posterior distribution not be strongly influenced by prior information.

As Θ and Φ are probabilities with limits 0 and 1, a natural choice of prior is the beta distribution: $\Theta \sim \text{Beta}(a_{\theta}, b_{\theta})$ and $\Phi \sim \text{Beta}(a_{\varphi}, b_{\varphi})$. The kernel of the joint prior distribution for (P, Θ, Φ) will then be

$$\pi(P, \theta, \varphi) \propto \theta^{a_{\theta}-1} (1-\theta)^{b_{\theta}-1} \varphi^{a_{\varphi}-1} (1-\varphi)^{b_{\varphi}-1} \mathbb{I}_{\{0 < 1, 0 < \theta < 1, 0 < \varphi < 1\}}.$$

We select the values of the hyperparameters so that the prior means and variances of Θ and Φ are equal to the point estimates and squared standard errors provided by a previous study:

$$a_{\theta} = \hat{\theta} \left(\frac{\hat{\theta}(1 - \hat{\theta})}{\hat{\sigma}_{\theta}^2} - 1 \right), \quad b_{\theta} = \left(\frac{(1 - \hat{\theta})}{\hat{\theta}} \right) a_{\theta}, \quad a_{\varphi} = \hat{\varphi} \left(\frac{\hat{\varphi}(1 - \hat{\varphi})}{\hat{\sigma}_{\varphi}^2} - 1 \right), \quad b_{\varphi} = \left(\frac{(1 - \hat{\varphi})}{\hat{\varphi}} \right) a_{\varphi}.$$

An alternative is to place independent normal prior distributions on $\Psi_{\theta} = \text{logit}(\Theta) \sim N(\tilde{\mu}_{\theta}, \tilde{\sigma}_{\theta}^2)$, the log odds of testing positive for a true positive and $\Psi_{\varphi} = \text{logit}(\Phi) \sim N(\tilde{\mu}_{\varphi}, \tilde{\sigma}_{\varphi}^2)$, the log odds of testing negative for a true negative, where $\text{logit}(x) = \log[x/(1 - x)]$. This choice is motivated by the fact that the maximum likelihood estimates of Ψ_{θ} and Ψ_{φ} should be approximately normally distributed for large enough samples. The priors for θ and φ will follow independent logistic normal distributions, and the kernel of the joint prior will be

$$\pi(P, \Theta, \Phi) \propto \frac{1}{\Theta(1 - \Theta)\Phi(1 - \Phi)} \exp \left[-\frac{1}{2} \left(\frac{(\text{logit}(\Theta) - \tilde{\mu}_{\theta})^2}{\tilde{\sigma}_{\theta}^2} + \frac{(\text{logit}(\Phi) - \tilde{\mu}_{\varphi})^2}{\tilde{\sigma}_{\varphi}^2} \right) \right] \mathbb{I}_{\left\{ 0 < P < 1, 0 < \Theta < 1, 0 < \Phi < 1 \right\}}.$$

When estimates of the sensitivity and specificity ($\hat{\theta}$ and $\hat{\varphi}$) and their standard errors ($\hat{\sigma}_{\theta}$ and $\hat{\sigma}_{\varphi}$) are available from previous studies, it is straightforward to obtain estimates of θ and φ via the delta method. We choose the hyperparameters for the prior distribution in Equation (10) accordingly:

$$\tilde{\mu}_{\theta} = \text{logit}(\hat{\theta}), \quad \tilde{\sigma}_{\theta}^2 = \hat{\sigma}_{\theta}^2 (\hat{\theta}(1 - \hat{\theta}))^{-2}, \quad \tilde{\mu}_{\varphi} = \text{logit}(\hat{\varphi}), \quad \tilde{\sigma}_{\varphi}^2 = \hat{\sigma}_{\varphi}^2 (\hat{\varphi}(1 - \hat{\varphi}))^{-2}.$$

3. Performance and comparison of estimators

In Section 3.1, we compare the large-sample performance of the MLEs $\hat{p}_1$ and $\hat{p}_2$ by calculating their asymptotic relative efficiency (ARE). This allows us to quantify the efficiency gain achieved by incorporating information on the number of test negatives. Intuitively, in situations where there are few false positives, performing gold standard verification among test positives should not noticeably improve estimation of the prevalence relative to estimates based only on the number of test positives and test negatives. Thus, in Section 3.2 we quantify the efficiency gain from verifying disease status among test positives by calculating the ARE of $\hat{p}_1$ relative to the estimator of Rogan and Gladen [12], $\hat{p}_R$. We do not show the ARE of $\hat{p}_2$ relative to $\hat{p}_R$ as we do not expect researchers to face a choice between these two estimators. This is because, when only the numbers of test negatives and test positives are known ($\hat{p}_R$ can be calculated), gold standard verification among test positives permits use of $\hat{p}_1$, so that $\hat{p}_2$ is not needed. Similarly, when only the

numbers of true and false positives are known ($\hat{p}_2$ can be calculated), obtaining the number of test negatives again permits use of $\hat{p}_1$, so that $\hat{p}_R$ is not needed. In Section 3.3, we present the results of a simulation study.

3.1. Efficiency gain from accounting for the number of test negatives

The MLE $\hat{p}_1$ assumes that the total number of individuals tested ($n = N_{00} + N_{01} + N_{11}$) is fixed, whereas the MLE $\hat{p}_2$ assumes that the number of test positives ($m = N_{01} + N_{11}$) is fixed while the total number tested ($N = m + N_{00}$) is random (and unknown). In order to compare the performance of $\hat{p}_1$ and $\hat{p}_2$, a common fixed sample size is required. We therefore slightly alter Scenario 1 in the following sense. Fixing m and p , we model the number of test negatives N_{00} using a negative binomial (NB) distribution with parameters m (the number of successes [test positives] required) and $p_z = \theta p + (1 - \varphi)(1 - p)$ (the probability of success [testing positive]). It can be shown that the point estimate $\hat{p}_1$ remains the same under this slightly altered model, but its asymptotic variance differs; one simply replaces the fixed total number tested n in Equation 5 with the expectation of the total number tested N under the NB model, $E[N] = m[\theta p + (1 - \varphi)(1 - p)]^{-1}$. Let σ_1^2 be this modified asymptotic variance. The relevant calculations are shown in the online supporting material, Section 2.1.

Using the variance of $\hat{p}_1$ under the altered Scenario 1, the asymptotic efficiency of $\hat{p}_2$ relative to $\hat{p}_1$ is

$$ARE(\hat{p}_1, \hat{p}_2) = \frac{\sigma_2^2}{\sigma_1^2} = 1 + \frac{p(1-p)(\varphi + \theta - 1)^2}{\theta(1-\varphi)(\varphi(1-p) + p(1-\theta))}.$$

The second term in this expression is non-negative for $0 < p < 1$, hence $\hat{p}_1$ is asymptotically more efficient than $\hat{p}_2$. We may interpret the ARE as follows. For large m , we would need to sample $ARE(\hat{p}_1, \hat{p}_2) \times m$ test positives in order to estimate p by $\hat{p}_2$ with the same precision as we would have achieved had we sampled m test positives and estimated p by $\hat{p}_1$. Equivalently, $ARE(\hat{p}_1, \hat{p}_2)$ represents the relative increase in the number of test positives that must be sampled in order to estimate p to a particular degree of precision when we discard information on the number of test negatives.

The magnitude of $ARE(\hat{p}_1, \hat{p}_2)$ depends on p , θ and φ . In particular, it is easy to see from Equation 12 $ARE(\hat{p}_1, \hat{p}_2) \rightarrow \infty$ as $\varphi \rightarrow 1$, but approaches a constant as $\theta \rightarrow 1$. $ARE(\hat{p}_1, \hat{p}_2)$ approaches a constant as $\varphi \rightarrow 0$ and approaches infinity as $\theta \rightarrow 0$, but these cases are not relevant in practice since θ and φ should be greater than 0.5 for any reasonable test. Consequently, the efficiency gain for p_1 relative to p_2 is most sensitive to changes in φ for meaningful values of θ and p . Figure 1(a) shows $ARE(\hat{p}_1, \hat{p}_2)$ as a function of φ for a range of values of p and θ , illustrating that under a range of reasonable assumptions ($\theta, p > 0.6$ and

0.4), $ARE(\hat{p}_1, \hat{p}_2)$ increases as each of p , θ and φ increases, and the magnitude of this increase is more sensitive to changes in φ than in θ . Plots of $ARE(\hat{p}_1, \hat{p}_2)$ as a function of θ for fixed values of p and φ are shown in the online supporting material, Section 2.3.

Figure 1(a) also shows that the asymptotic efficiency gain from including information on test negatives is greatest for high prevalence diseases with accurate screening tests (particularly highly specific tests), and is least for low prevalence diseases with less accurate screening tests. This result agrees with intuition in that including information on test negatives allows us to compute the probability of testing positive, p_z . For high prevalence diseases and/or accurate screening tests, p_z will be closer to p , and so we stand to gain more information on the true prevalence in such cases.

3.2. Efficiency gain from verifying disease status in test positives

We now present the ARE of $\hat{p}_R$ (Equation 1), an estimator of p that uses information on the number of test negatives and test positives only, compared to $\hat{p}_1$, which additionally incorporates information on true disease status among test positives. The derivation is shown in the online supporting material, Section 2.2. We find

$$ARE(\hat{p}_1, \hat{p}_R) = \frac{\sigma_2^2}{\sigma_1^2} = 1 + \frac{\theta(1-\varphi)((1-\theta)p + \varphi(1-p))}{p(1-p)(1-\theta-\varphi)^2}.$$

We see that $ARE(\hat{p}_1, \hat{p}_2) > 1$ for all $\theta, \varphi, p \in (0,1)$.

Figure 1(b) shows $ARE(\hat{p}_1, \hat{p}_R)$ as a function of φ for fixed values of θ and p , demonstrating that $\hat{p}_1$ is typically much more asymptotically efficient than $\hat{p}_R$, except when θ is close to 1. Intuitively, when the specificity is high, there will be few false positives in the dataset and so $\hat{p}_1$ uses little additional information relative to $\hat{p}_R$. Notice also that $ARE(\hat{p}_1, \hat{p}_R) - 1 = 1/(ARE(\hat{p}_1, \hat{p}_2) - 1)$, so that whenever the efficiency of $\hat{p}_1$ increases relative to $\hat{p}_2$, it decreases relative to $\hat{p}_R$, and vice versa. This result also agrees with intuition, since $\hat{p}_R$ uses data that inform solely about p_z , while $\hat{p}_2$ uses no information on p_z . Plots of $ARE(\hat{p}_1, \hat{p}_R)$ as a function of θ for fixed values of p and φ are shown in the online supporting material, Section 2.3.

3.3. Simulation study

We conducted a simulation study to compare the bias and standard error (s.e.) of the estimators $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_R$, and the coverage of their CIs. We ran 100,000 simulations for each of 27 different combinations of the true prevalence p ($p = 0.01, 0.1$ or 0.4), fixed number of test positives ($m = 1000, 500$ or 100), and true sensitivity and specificity ($\theta = \varphi = 0.6, 0.9$ or 0.99). To compute $\hat{p}_1$ and $\hat{p}_R$, the number of test negatives was simulated from a NB

distribution as described in Section 3.1. We first assumed that the true θ and ϕ are known and the correct values are used when calculating the prevalence estimates.

Figure 2 shows the mean and 2.5th and 97.5th empirical percentiles of $\hat{p}_1$ and $\hat{p}_2$ from each set of simulations. Table 2 shows the coverage of the 95% CIs as estimated by the proportion of simulations for which the 95% CI contained the true p . Full results, including those for $\hat{p}_R$, are shown in the online supporting material, Section 3.

Figure 2 suggests that $\hat{p}_1$ and $\hat{p}_2$ have comparable variance and are relatively unbiased for low prevalence diseases ($p = 0.1$) and less accurate tests ($\theta, \phi = 0.9$). However, for high prevalence diseases and very accurate screening tests, the variance of $\hat{p}_2$ was much greater than that of $\hat{p}_1$. Additional simulations (not shown) in which one of θ and ϕ was fixed at 0.9 and the other varied suggested that this high variance was driven primarily by high specificity rather than sensitivity.

Table 2 shows that the coverage of the asymptotic CIs for both estimators was close to the nominal level of 95% in most simulations. However, for low prevalence diseases and low accuracy tests, the coverage rates were poor for both estimators when the number of test positives was relatively small. Additional simulations suggested that this was driven by both low sensitivity and low specificity. The estimated coverage of the CI for $\hat{p}_2$ was also poor for high prevalence diseases with very accurate tests; additional simulations suggested that this was driven by high specificity ($\phi = 0.99$).

Next, we considered misspecification of θ and ϕ . We simulated a plausible real-world situation wherein previous estimates of the sensitivity and specificity of the rapid test, θ^* and ϕ^* , were obtained in a reference laboratory under ideal conditions, while the true θ and ϕ as applied in clinical practice were lower due to imperfect administration of the test. This was achieved by assuming that the test yielded 10% fewer false positives and false negatives when administered in a reference laboratory compared to clinical practice, i.e., $(1 - \theta^*) = 0.9(1 - \theta)$ and $(1 - \phi^*) = 0.9(1 - \phi)$. This leads to a misspecified sensitivity of $\theta^* = 0.64$ when the truth is $\theta = 0.6$, $\theta^* = 0.91$ when $\theta = 0.9$, $\theta^* = 0.991$ when $\theta = 0.99$, and similarly for the specificity.

Misspecification of the sensitivity and specificity had relatively little impact on the Scenario 1 estimator for accurate tests ($\phi, \theta = 0.9$), leading to only small changes in bias (Figure 2 and online supporting material Table 2) and reductions in the coverage of the CIs (Table 2) compared to correct specification of θ and ϕ . More important bias of between roughly -0.03 and -0.01 and poorer coverage were observed when the true sensitivity and specificity were low ($\theta, \phi = 0.6$) and the true prevalence was high ($p = 0.1$ or 0.4). Misspecification had a greater impact on the Scenario 2 estimator. Under the misspecification scenarios considered, coverage of the CIs for $\hat{p}_2$ was poor, and increases in the bias were seen for most values of θ and ϕ . This bias was greatest for low accuracy tests or high true prevalences, and was relatively small for high accuracy tests and low true prevalences.

4. Illustrative examples

In this section, we illustrate the application of the prevalence estimators introduced in this paper using data from two screening studies in Swaziland. In the first example, the number of test negatives is known, so that the Scenario 1 estimator may be used. For the second example, the number of test negatives is unknown, and only the Scenario 2 estimator can be applied. Table 3 summarizes the results from the two data examples, and also shows results using the estimator of Rogan and Gladen [12] for comparison.

4.1. Example 1: Known number of test negatives

The first example uses data from a nation-wide study, the Swaziland HIV Incidence Measurement Survey (SHIMS). A nationally representative sample of participants aged 18—49 was recruited via a two-stage household sampling scheme between December 2010 and July 2011 [13]. Here, we are accessing a publicly-available data set, which represents a randomly-selected subset of 5% of the total study population [14]. For the purposes of this illustrative example, we assume that the individuals in this random subset are independent. HIV testing was performed using the Determine®HIV-1/2 Ag/Ab Combo rapid test, with confirmatory tests performed under some circumstances. For this analysis, we assume that only the Determine rapid test was used, since the number for whom secondary testing was performed was small (less than 1% of test positives). Among those who tested positive, HIV status was verified by viral load measurement. Those with viral load below the limit of detection were considered to be false positives. We excluded participants who reported previous exposure to ART.

In the available data, $N_0 = 607$ participants tested negative and $N_1 = 209$ tested positive; of these, 9 were found to be false positives. A previous study estimated the sensitivity and specificity of the rapid test to be 0.994 (s.e. 0.006) and 0.992 (s.e. 0.003) respectively [15]. We therefore set $\theta = 0.994$ and $\phi = 0.992$. Using the frequentist Scenario 1 estimator and asymptotic CI presented in Section 2.1.1, the estimated prevalence was 0.247 with 95% CI (0.217, 0.276). Using the Scenario 2 estimator and CI of Section 2.1.2, the estimated prevalence was 0.147 with 95% CI (0.063, 0.231).

Next, we applied the Bayesian estimation procedure described in Section 2.2. We used both the Scenario 1 (Equation 2) and Scenario 2 (Equation 6) likelihoods first with logistic normal priors and then with beta priors for the sensitivity and specificity. The hyperparameters of the logistic normal prior were taken to be $\tilde{\mu}_\theta = 5.09$, $\tilde{\sigma}_\theta^2 = 1.00$, $\tilde{\mu}_\phi = 4.86$ and $\tilde{\sigma}_\phi^2 = 0.20$. The hyperparameters of the beta prior were $a_\theta = 161.01$, $b_\theta = 0.99$, $a_\phi = 642.0$ and $b_\phi = 4.99$. Samples were drawn from the posterior using Stan via the Rstan package in R. We ran ten chains for 2,000 scans each, discarding the first 500 scans from each chain as the burn-in period. For both Bayesian models, the potential scale reduction factors (R) from the MCMC chains were in the range 1.000 to 1.002 for all parameters, indicating convergence. Trace plots of the posterior samples for all Bayesian analyses are shown in Section 4 of the online supporting material.

With the Scenario 1 estimator, we estimated the posterior mean for P to be 0.248 with 95% credible interval (0.219, 0.279) using the logistic normal prior and 0.248 (0.218, 0.279) using the beta prior—scarcely different from the frequentist results. With the Scenario 2 estimator, we estimated the posterior mean for P to be 0.195 with 95% credible interval (0.071, 0.394) using the logistic normal prior and 0.181 (0.062, 0.365) using the beta prior. The credible intervals are wider than the frequentist confidence intervals for both estimators, although this effect is much more pronounced for Scenario 2. This indicates that the frequentist approach may underestimate the variability of the prevalence estimate when there is uncertainty in the sensitivity and specificity, especially when information on the number of test negatives is not available. The Scenario 2 estimator also appears to be more sensitive to the choice of prior likely due to poorer identifiability in this model, although this has no material impact on interpretation in light of the wide credible intervals.

Table 3 summarizes the frequentist and Bayesian prevalence estimates. For both approaches, the Scenario 2 estimate is much lower than the Scenario 1 estimate, although the confidence/credible intervals do overlap. In high prevalence, high accuracy settings with a comparatively small number of test positives (< 500) such as this one, our simulation results from Section 3.3 showed that $\hat{p}_2$ is subject to large sampling variability. This could partly explain the observed discrepancy between the two estimators. The simulations also showed that in cases where θ and ϕ are misspecified at values greater than the truth, the Scenario 2 estimator may be biased low. This situation is plausible here: the assumed θ and ϕ were estimated in an idealized laboratory setting, while the true sensitivity and specificity in practice may be lower. In sum, the Scenario 1 estimates are much more reliable.

Table 3 also shows the prior and posterior means and credible intervals for Θ and Φ . The prior means and 2.5th/97.5th percentiles were close to their posterior counterparts for both the Scenario 1 and 2 estimators. However, there appears to be a small amount of learning in $\$$ for the Scenario 1 estimator. For example, using the logistic normal prior, the prior mean (2.5th, 97.5th percentile) of $\$$ was 0.991 (0.982, 0.997) compared to the posterior mean of 0.988 (0.981, 0.994). The posterior means for θ in both Scenario 1 and Scenario 2 and for ϕ in Scenario 2 changed by 0.001 or less compared to the relevant prior means. Figures 3 (a) and (b) show histograms of the posterior samples along with scatter plots illustrating the joint distribution of each pair of parameters for the Scenario 1 estimator.

4.2. Example 2: Unknown number of test negatives

In the second example, we analysed preliminary data from the Maximizing ART for Better Health and Zero New HIV Infections (MaxART) study, conducted at 14 health facilities in Swaziland. Participants were new, ART-naïve clients enrolling in the 14 study clinics between August 2014 and February 2016. Pregnant and breastfeeding women were excluded. HIV screening was performed using the Determine® HIV-1/2 rapid test. For those who tested positive, HIV status was verified by viral load measurement. Those with viral load below the limit of detection were considered to be false positives. The number of test negatives was not recorded in this study since only test positives were recruited, so that only the Scenario 2 estimator applies here.

During the study period, $N_i = 1778$ clients tested positive, of whom $N_{01} = 140$ were found to be false positives. A previous study estimated the sensitivity and specificity of the rapid test used here to be 1.000 and 0.996 [16]. We assumed $\theta = 0.997$ (s.e. = 0.003) and $\varphi = 0.996$ (s.e. = 0.002), where the former was obtained by calculating the proportion of true positives who tested positive in [16] after adding one to each cell, in order to avoid assuming $\theta = 1$. This was done primarily for consistency with the Bayesian analysis presented below, although $\theta = 1$ is an acceptable assumption for the frequentist estimator and small changes in θ made relatively little difference to the estimate. Using the frequentist Scenario 2 estimator and asymptotic confidence interval presented in Section 2.1.2, the estimated prevalence was $\hat{p}_2 = 0.043$ with 95% CI 0.036 to 0.050.

Next, we carried out Bayesian prevalence estimation using the Scenario 2 likelihood (Equation 6). The hyperparameters of the logistic normal prior were be $\tilde{\mu}_\theta = 5.97$, $\tilde{\sigma}_\theta^2 = 1.00$, $\tilde{\mu}_\varphi = 5.56$ and $\tilde{\sigma}_\varphi^2 = 0.25$. For the beta prior, we set $a_\theta = 390.00$, $b_\theta = 0.99$, $a_\varphi = 1038.00$ and $b_\varphi = 3.98$. Using the same sampling procedure as in Section 4.1, we estimated the posterior mean for P (95% credible interval) to be 0.059 (0.021, 0.126) using the logistic normal prior and 0.052 (0.018, 0.103) using the beta prior. Note that the credible intervals are much wider than the frequentist CI, again highlighting the importance of accounting for uncertainty in θ and φ in Scenario 2. For both Bayesian models, $\hat{R}$ from the MCMC chains was in the range 1.001 to 1.002 for all parameters, indicating that the chains had converged. Figures 3 (c) and (d) show histograms of the posterior samples along with scatter plots illustrating the joint distributions of pairs of parameters. Table 3 shows the prior and posterior means and 95% credible intervals for Θ and Φ . The posteriors for both Θ and Φ appear to be very close to their respective priors, indicating little learning in these parameters.

5. Discussion

We have developed frequentist and Bayesian approaches for estimating HIV prevalence from screening data when HIV status is verified by a gold standard (e.g., viral load or Western blot) among test positives only, and the sensitivity and specificity of the test are known. Data of this type are routinely collected in a range of clinical settings, not limited to HIV testing, since it is common to seek verification of a diagnosis following a positive result on an imperfect screening test. When applied in this context, this method provides an estimate of disease prevalence among people eligible for screening.

These estimators can be applied when the number of test negatives is known or unknown. Under some circumstances, incorporating information about the number of test negatives (Scenario 1) leads to substantial improvements in the bias and precision of prevalence estimates. Our simulation study indicated that this advantage is greatest for accurate tests—most importantly, tests with high specificity—especially when the true disease prevalence in the target population is high.

In particular, for true prevalences greater than 10% and sensitivity and specificity greater than 90%, the estimator that did not incorporate information on test negatives (Scenario 2) had a large variance in simulations unless the number of test positives sampled was 1000 or

greater, while the Scenario 1 estimator still performed well. Indeed, our analytic results show that for specific tests ($\phi > 0.9$) applied to prevalent diseases ($p > 0.2$), the variance of the Scenario 2 estimator was more than double that of the Scenario 1 estimator in large samples. In the most extreme case when there are no false positives in the data (a potentially common situation in small samples with highly specific tests), the Scenario 2 estimator will be equal to one with CI (1,1)—not very informative. Meanwhile, the Scenario 1 estimator and its CI will still be useful (different from one, with non-zero interval width). However, note that under these circumstances the Scenario 1 estimator will typically be close to the estimator of Rogan and Gladen, which requires no gold standard verification; in fact, they are equal when there are no false positives in the data and $\phi = 1$ (calculations not shown).

On the other hand, for less accurate tests (sensitivity and specificity below 90%), the Scenario 2 estimator performed adequately and was often very similar to the Scenario 1 estimator. For routinely collected data, particularly in resource-limited settings, the number of test negatives is usually not recorded. In such circumstances, our results suggest that it is still possible to estimate disease prevalence with reasonable accuracy. Nevertheless, we recommend recording the number of test negatives whenever there is interest in estimating disease prevalence from screening data of this type, as an inexpensive way of improving the quality of the estimate. Notably, however, for low prevalence ($< 10\%$) diseases, the coverage of the CIs presented for both the Scenario 1 and the Scenario 2 estimators was relatively poor and below the nominal level of 95% in simulations, even when 500 or more test positives were sampled.

The Scenario 1 estimator will be similar to the proportion of test positives when sensitivity and specificity are high and/or prevalence is high. As progress is made toward the first of the 90–90–90 goals—to ensure that 90% of all HIV positive people know their HIV status—the prevalence of HIV in the screening target population will drop. When this happens, the bias in the proportion of test positives as an estimator of the true prevalence will increase [12]. The Scenario 1 estimator ($\hat{p}_1$) represents a cheap and effective method of reducing this bias. While the estimator of Rogan and Gladen ($\hat{p}_R$), which is based only on the proportion of test positives but is corrected for imperfect sensitivity and specificity, will also reduce this bias, $\hat{p}_1$ is more efficient and performs better than $\hat{p}_R$ under a range of circumstances. In practice, the benefits of estimating p by $\hat{p}_1$ relative to $\hat{p}_R$ may need to be weighed against the cost of gold standard disease verification among test positives.

A limitation of the frequentist prevalence estimators described in this paper is the requirement that the sensitivity and specificity of the tests be known. In HIV testing, while the sensitivity and specificity of most screening tests have been measured accurately in reference laboratories under ideal circumstances, they may be lower in practice due to, for example, unplanned deviations from the testing protocol. When performing a Bayesian analysis, one approach for allowing for such deviations would be to alter the prior distributions on the sensitivity and specificity accordingly. Alternatively, one could vary the assumptions used to obtain frequentist estimates in a sensitivity analysis. Our simulation results suggest that the prevalence estimate is particularly sensitive to misspecification of the sensitivity and specificity when the number of test negatives is unknown. Under these

circumstances, the Bayesian approach is especially recommended, as it will give credible intervals that appropriately account for the uncertainty in the sensitivity and specificity.

In the MaxART Swaziland study where the number of test negatives was unknown, we estimated HIV prevalence to be 5.9%, with 95% credible interval 2.1% to 12.6%. This compares with 24.8% (95% credible interval 21.9% to 27.9%) as estimated using publicly-available data from the SHIMS study, where information on test negatives was available. The target populations of the two studies differed in several important respects. SHIMS aimed to recruit a nationally representative sample of adults in Swaziland from 2010 to 2011 using standard sample survey methods, while MaxART recruited adults presenting for care in 14 specific clinics from 2014 to 2016. When using the methods presented in this paper, investigators should take care to specify the population to whom the results can be generalized. In the case of MaxART, although the estimate may not be nationally representative, it likely captures the prevalence among new ART-naïve clients at the study clinics—information that is nonetheless important for national programs, projections and planning. The results from SHIMS, meanwhile, estimate the national prevalence among ART-naïve adults from 2010 to 2011, several years before MaxART.

A possible explanation for the difference in the two estimates, then, is that the majority of HIV positive people eligible for screening at the MaxART clinics already had a diagnosis with relatively few new infections occurring, such that the prevalence among the rest of those eligible for screening was low compared to national levels. We also note that HIV prevalence in Swaziland is likely declining over time, as it is throughout Sub-Saharan Africa. Alternatively, it is possible that the sensitivity and specificity of the rapid test used in MaxART are lower than estimated in reference laboratories, which may contribute to underestimation of the prevalence. Gold standard verification of HIV status in (possibly a subsample of) test negatives could help to distinguish between these possibilities by allowing (1) direct estimation of the sensitivity and specificity of the screening algorithm as applied in the field, and (2) more efficient estimation of HIV prevalence among those currently targeted for screening. The prevalence estimators presented in this paper will be useful in studies and screening programs where gold standard verification of test positives is already a part of the protocol or when adding gold standard verification to the protocol in at least a subset of test positives is an option going forward. Further investigation into the recruitment strategy will likely be required if the estimated prevalence is different than expected.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. United Nations Programme on HIV/AIDS. 90–90–90: an ambitious treatment target to help end the AIDS epidemic. Technical Report, United Nations, Geneva 2014.
2. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *The Lancet* 2009; 373(9657):48–57.

3. WHO. Consolidated guidelines on HIV testing services. Technical Report, World Health Organization 2015.
4. on HIV/AIDS JUNP, et al. Fast track: ending the AIDS epidemic by 2030. Geneva: UNAIDS 2014;.
5. Walter SD. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. *Epidemiology* 1999; 10(1):67–72. [PubMed: 9888282]
6. Li F, Chu H, Nie L. A two-stage estimation for screening studies using two diagnostic tests with binary disease status verified in test positives only. *Statistical Methods in Medical Research* 2011; 0(0):1–22.
7. Chock C, Irwig L, Berry G, Glasziou P. Comparing dichotomous screening tests when individuals negative on both tests are not verified. *Journal of Clinical Epidemiology* 1997; 50(11):1211–1217. [PubMed: 9393377]
8. Berry G, Smith CL, Macaskill P, Irwig L. Analytic methods for comparing two dichotomous screening or diagnostic tests applied to two populations of differing disease prevalence when individuals negative on both tests are unverified. *Statistics in Medicine* 2002; 21(6):853–862. [PubMed: 11870821]
9. Stock EM, Stamey JD, Sankaranarayanan R, Young DM, Muwonge R, Arbyn M. Estimation of disease prevalence, true positive rate, and false positive rate of two screening tests when disease verification is applied on only screen-positives: a hierarchical model using multi-center data. *Cancer Epidemiology* 2012; 36(2):153–160. [PubMed: 21856264]
10. Liu J, Chen F, Yu H, Zeng P, Liu L. A two-stage Bayesian method for estimating accuracy and disease prevalence for two dependent dichotomous screening tests when the status of individuals who are negative on both tests is unverified. *BMC Medical Research Methodology* 2014; 14(1): 110. [PubMed: 25245321]
11. Chu H, Zhou Y, Cole SR, Ibrahim JG. On the estimation of disease prevalence by latent class models for screening studies using two screening tests with categorical disease status verified in test positives only. *Statistics in Medicine* 2010; 29(11):1206–1218. [PubMed: 20191614]
12. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology* 1978; 107(1):71–76. [PubMed: 623091]
13. Justman J, Reed JB, Bicego G, Donnell D, Li K, Bock N, Koler A, Philip NM, Mlambo CK, Parekh BS, et al. Swaziland HIV incidence measurement survey (SHIMS): a prospective national cohort study. *The Lancet HIV* 2017; 4(2):e83—e92. [PubMed: 27863998]
14. ICAP. Swaziland HIV incidence measurement survey 2016 URL <http://shims.icap.columbia.edu/>.
15. Rosenberg NE, Kamanga G, Phiri S, Nsona D, Pettifor A, Rutstein SE, Kamwendo D, Hoffman IF, Keating M, Brown LB, et al. Detection of acute HIV infection: a field evaluation of the Determine® HIV-1/2 Ag/Ab combo test. *Journal of Infectious Diseases* 2012; 205(4):528–534. [PubMed: 22207651]
16. Lyamuya EF, Aboud S, Urassa WK, Sufi J, Mbwana J, Ndugulile F, Massambu C. Evaluation of simple rapid HIV assays and development of national rapid HIV test algorithms in Dar es Salaam, Tanzania. *BMC Infectious Diseases* 2009; 9(1):19. [PubMed: 19226452]

Figure 1.

(a) $ARE(\hat{p}_1, \hat{p}_2)$ and (b) $ARE(\hat{p}_1, \hat{p}_R)$ as a function of specificity (ϕ) for different values of sensitivity (θ) and prevalence (p).

Figure 2.

$\hat{p}_1$ and $\hat{p}_2$ computed from simulated data when the sensitivity and specificity are correctly specified, i.e., $\theta^* = \theta$ and $\varphi^* = \varphi$, or when they are misspecified as $\theta^* = 0.9\theta + 0.1$ and $\varphi^* = 0.9\varphi + 0.1$, where θ , φ are the true values and θ^* , φ^* are the assumed values used in calculating the prevalence estimates. Points represent the mean from 100,000 simulations and bars range from the 2.5th to the 97.5th percentiles. Results are displayed with different X axis scales for simulations conducted using different values of the true prevalence p ; the vertical dotted line in each plot represents this true value.

(a) SHIMS data example, Scenario 1, logistic normal prior

(b) SHIMS data example, Scenario 1, beta prior

(c) MaxART data example, Scenario 2, logistic normal prior

(d) MaxART data example, Scenario 2, beta prior

Figure 3.

Histograms and heatmaps showing samples drawn from the posterior distribution of P , Θ and Φ for the data examples in Sections 4.1 and 4.2.

Table 1.

Hypothetical study data

		Gold standard X		
		Positive ($X = 1$)	Negative ($X = 0$)	Total
Screening test Z	Positive ($Z = 1$)	N_{11}	N_{01}	$N_{\cdot 1}$ (m when fixed [*])
	Negative ($Z = 0$)	N_{10}	N_{00}	$N_{\cdot 0}$
	Total	$N_{1\cdot}$	$N_{0\cdot}$	N (n when fixed [*])

* Lower case letters denote fixed quantities, and upper case letters denote random variables. Quantities may be random in some sections of this paper and fixed in others, as indicated at the start of the relevant sections.

Table 2.

Estimated coverage of the 95% CIs for $\hat{p}_1, \hat{p}_2$ from 100,000 simulations.

		Scenario 1 estimator, $\hat{p}_1$						Scenario 2 estimator, $\hat{p}_2$					
		θ, ϕ correctly specified ^a			θ, ϕ misspecified ^b			θ, ϕ correctly specified ^a			θ, ϕ misspecified ^b		
p	m	True θ, ϕ	0.6	0.9	0.99	0.6	0.9	0.99	0.6	0.9	0.99	0.6	0.9
0.01	1000	0.932	0.948	0.949	0.950	0.890	0.943	0.949	0.919	0.952	0.951	0.811	0.779
0.01	500	0.933	0.945	0.949	0.950	0.872	0.941	0.949	0.932	0.938	0.952	0.864	0.839
0.01	100	0.778	0.925	0.947	0.949	0.779	0.922	0.949	0.778	0.922	0.948	0.779	0.846
0.10	1000	0.947	0.950	0.950	0.950	0.753	0.937	0.949	0.950	0.949	0.952	0.471	0.511
0.10	500	0.947	0.949	0.950	0.950	0.841	0.941	0.951	0.949	0.948	0.949	0.688	0.865
0.10	100	0.936	0.948	0.951	0.951	0.900	0.944	0.950	0.945	0.945	0.927	0.861	0.863
0.40	1000	0.949	0.951	0.950	0.950	0.388	0.926	0.950	0.950	0.953	0.951	0.234	0.726
0.40	500	0.949	0.950	0.950	0.950	0.641	0.937	0.949	0.950	0.953	0.939	0.519	0.817
0.40	100	0.946	0.948	0.950	0.950	0.878	0.944	0.947	0.954	0.937	0.715	0.861	0.919

^aResults for simulations with correctly specified sensitivity and specificity, i.e., $\theta^* = \theta$ and $\phi^* = \phi$ where θ, ϕ are the true values and θ^*, ϕ^* are the assumed values used in calculating the prevalence estimates.

^bResults for simulations with misspecified sensitivity and specificity, where $\theta^* = 0.9\theta + 0.1$ and $\phi^* = 0.9\phi + 0.1$.

Table 3.

Estimated HIV prevalence in two studies in Swaziland

Data source	Estimator	Prior mean (95% credible interval)			Estimate (95% interval) ^a		
		θ	φ	p	θ	φ	p
SHIMS							
Scenario 1							
	Frequentist						
	$\widehat{p}_1$	0.994 ^b	0.992 ^b	—	—	—	0.247 (0.217, 0.276)
	$\widehat{p}_R^{[12]}$	0.994 ^b	0.992 ^b	—	—	—	0.252 (0.251, 0.253)
	Bayesian						
	Logistic normal prior	0.990 (0.958, 0.999)	0.991 (0.982, 0.997)	0.500 (0.025, 0.975)	0.990 (0.959, 0.999)	0.988 (0.981, 0.994)	0.248 (0.219, 0.279)
	Beta prior	0.994 (0.977, 1.000)	0.992 (0.984, 0.997)	0.500 (0.025, 0.975)	0.994 (0.977, 1.000)	0.989 (0.983, 0.994)	0.248 (0.218, 0.279)
Scenario 2							
	Frequentist						
	$\widehat{p}_2$	0.994 ^b	0.992 ^b	—	—	—	0.147 (0.063,0.231)
	Bayesian						
	Logistic normal prior	0.990 (0.958, 0.999)	0.991 (0.982, 0.997)	0.500 (0.025, 0.975)	0.990 (0.956, 0.999)	0.990 (0.980, 0.996)	0.195 (0.071, 0.398)
	Beta prior	0.994 (0.977, 1.000)	0.992 (0.984, 0.997)	0.500 (0.025, 0.975)	0.994 (0.978, 1.000)	0.991 (0.983, 0.997)	0.181 (0.062, 0.365)
MaxART							
Scenario 2							
	Frequentist						
	$\widehat{p}_2$	0.997 ^b	0.996 ^b	—	—	—	0.043 (0.036, 0.050)
	Bayesian						
	Logistic normal prior	0.996 (0.982, 1.000)	0.996 (0.990, 0.999)	0.500 (0.025, 0.975)	0.996 (0.982, 1.000)	0.995 (0.988, 0.998)	0.059 (0.021, 0.126)
	Beta prior	0.997 (0.991, 1.000)	0.996 (0.992, 0.999)	0.500 (0.025, 0.975)	0.997 (0.991, 1.000)	0.995 (0.991, 0.998)	0.052 (0.018, 0.103)

^aFor frequentist estimates, we show the point estimate and 95% confidence interval. For Bayesian estimates, we show the posterior mean along with the 2.5th and 97.5th percentiles of the posterior distribution.

^bFor the frequentist estimators, we show the fixed input values of θ and φ used in computing the estimate of p