Published in final edited form as:

Stat Methods Med Res. 2016 April ; 25(2): 917–935. doi:10.1177/0962280212473514.

Estimating the Prevalence of Transmitted HIV Drug Resistance Using Pooled Samples

Mariel M. Finucane^{*,1}, Christopher F. Rowley^{2,3,4}, Christopher J. Paciorek⁵, Max Essex^{3,4}, and Marcello Pagano⁶

¹The J. David Gladstone Institutes, University of California, San Francisco, USA

²Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Boston, USA

³Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, USA

⁴Harvard School of Public Health AIDS Initiative, Boston, USA

⁵Department of Statistics, University of California, Berkeley, USA

⁶Department of Biostatistics, Harvard School of Public Health, Boston, USA

Abstract

In many resource-poor countries, HIV-infected patients receive a standardized antiretroviral (ART) cocktail. In these settings, population-level surveillance of drug resistance is needed to characterize the prevalence of resistance mutations and to enable ART programs to select the optimal regimen for their local population. The surveillance strategy currently recommended by the World Health Organization is prohibitively expensive in some settings and may not provide a sufficiently precise rendering of the emergence of drug resistance. By using a novel assay on pooled sera samples, we decrease surveillance costs while simultaneously increasing the accuracy of drug resistance prevalence estimates for an important mutation that impacts first-line antiretroviral therapy. We present a Bayesian model for pooled-testing data that garners more information from each resistance assay conducted, compared with individual testing. We expand on previous pooling methods to account for uncertainty about the population distribution of within-subject resistance levels. In addition, our model accounts for measurement error of the resistance assay, and this added uncertainty naturally propagates through the Bayesian model to our inference on the prevalence parameter. We conduct a simulation study that informs our pool size recommendations and that shows that this model renders the prevalence parameter identifiable in instances when an existing non-model-based estimator fails.

Keywords

pooling; HIV; primary drug resistance; Bayesian modelling; measurement error

*1650 Owens St; San Francisco, CA; 94117, mariel.finucane@gladstone.ucsf.edu, 415 · 734 · 2575.

Conflict of Interest Statement

The Authors declare that there is no conflict of interest.

1 Introduction

Antiretroviral therapy (ART) for HIV/AIDS is becoming increasingly available in resource-poor settings, turning what used to be a certain death sentence into a manageable chronic disease. The emergence of drug resistant HIV, however, has the potential to compromise deeply the benefits of ART programs¹.

In the United States, resistance assays are recommended at the time of HIV diagnosis and prior to initiation of therapy², allowing clinicians to tailor treatment regimens according to patients' resistance profiles. For those patients with drug-resistant virus, this individualized approach yields outcomes that are no worse than those obtained by non-resistant individuals³.

Unfortunately, drug resistance genotyping is too expensive to be available currently for clinical use in the developing world. Instead, in many resource-poor countries all patients receive a common first-line ART regimen. Under this public-health approach to ART scale-up, regimens are selected at the population level and not at the individual level as in the United States. When the first-line drugs fail, a standardized second-line regimen may be prescribed. Without knowledge of a given patient's resistance profile, however, there can be no certainty that either course of therapy will suppress viral replication successfully, particularly given the expected increase in circulating HIV drug resistance due to expanding ART use.

Without individual-level drug resistance profiling, population-level surveillance of drug resistance is essential. By characterizing the population-level prevalence of resistance mutations, surveillance enables ART programs to select optimal first- and second-line regimens, thus increasing the probability of ART success for a given population.

To this end, the World Health Organization (WHO) has begun to implement a surveillance system for classifying the prevalence of transmitted resistance as less than 5%, 5–15%, or over 15% using binomial sequential lot quality assurance sampling (BSLQAS)^{4,5}. Samples from up to 47 treatment-naïve individuals consecutively diagnosed with HIV are sequenced to allow this classification. Although this method is cost-effective, there are three important reasons why it may be suboptimal for drug resistance surveillance in the developing world.

Firstly, even the cost of 47 genotypic assays may be prohibitive in some settings. In Malawi, for example, there are 103 public ART clinics distributed throughout the entire country. In 2006, given cost constraints, only three antenatal care clinics, all located in the capital, were able to participate in the WHO surveillance program⁶. There is likely geographic variability in the prevalence of drug resistance, but without more affordable surveillance methods, this variability will not be measurable.

Secondly, the WHO surveillance program uses population sequencing, a method that provides information about all mutations in a 1.8-kb region of the HIV genome. However, since the goal of surveillance is to maximize the effectiveness of first-line regimens for a treatment-naïve individual infected with a drug-resistant virus, only a few specific point mutations may be relevant. Resources should be focused on those mutations most likely to be selected for by a country's standardized regimens, and thus most likely to be circulating. For this reason, most

of the information obtained by sequencing such a large portion of the genome may be extraneous.

In Botswana, for example, first-line therapy has historically consisted of zidovudine (AZT), lamivudine (3TC), and either efavirenz (EFV) or nevirapine (NVP). Recently, AZT and 3TC have been replaced by a fixed-dose combination of emtricitabine and tenofovir. In monitoring drug resistance to this first-line regimen, it may be more prudent to focus on the prevalence of population-wide resistance to these medications, rather than obtaining data regarding resistance to drugs that have never been used on a wide scale in Botswana. This approach will maximize scarce resources while still providing meaningful and useful data.

Lastly, the WHO surveillance program may not provide a sufficiently precise rendering of the emergence of drug resistance. Of the nine completed surveys that Bennett et al.⁵ describe, all found the prevalence of transmitted drug resistance to be less than 5%. More recently, Aghokeng et al.⁷ found the prevalence of transmitted drug resistance to be less than 5% in Chad. Vardavas and Blower⁸ argue that the prevalences in other African countries are likely to fall below this threshold as well, concluding that “although the BSLQAS method using a threshold value of 5% requires a small sample size and is relatively inexpensive it may actually not be cost-effective at the early stages of the treatment program.” In settings where prevalence is still very low, a low-cost sentinel surveillance system is needed to provide a higher-resolution picture of the emergence of primary drug resistance. Although Aghokeng et al.⁷ found the prevalence of transmitted drug resistance to be greater than 5% in Cameroon, they still focus on the challenges presented by the WHO methodology, including cost and workload in particular.

The aim of this paper is to specify a novel statistical method for analyzing results from a new laboratory assay⁹, to allow detection of transmitted resistance at lower prevalence values than the WHO method and at a lower cost. We pursue prevalence estimation rather than classification, allowing us to provide the earliest possible warning of emerging primary drug resistance under new public-sector treatment programs, in settings where the population prevalence of resistant virus is still likely to be less than 5%.

The idea underlying our method is to cut costs by pooling blood samples from multiple individuals and performing a single allele-specific polymerase chain reaction (ASPCR) assay on the pool. Dorfman¹⁰ first introduced pooled methods with the goal of “detection of defective members of large populations”. Pooled detection algorithms have since been tailored for HIV research, optimizing screening for acute infections^{11;12} and, more recently, for virologic failure^{13;14}. Rather than individual detection though, our goal is prevalence estimation, building on the work of Tu et al.¹⁵. Their method was adapted in Wein and Zenios¹⁶ and in Zenios and Wein¹⁷ to allow inference based on a continuous assay result.

In modeling drug resistance, an important challenge is that within an infected individual, HIV exists as a viral quasispecies, a group of genetically diverse strains. For this reason, an individual with drug resistance also harbors some non-zero proportion of non-resistant virus. A limitation of previous work^{17;18;19} is the requirement that the population distribution of these within-subject resistance levels be known. Our model relaxes this assumption, using

the Bayesian framework to account explicitly for our uncertainty about the parameters of this distribution.

Another challenge that arises in the pooled analysis of drug resistance data is that combining samples from subjects with differing viral loads may lead some resistance profiles to be under- or over-represented in the pooled sample. In order to avoid a situation where a sample with a high viral load of wild-type HIV mitigates the detection of a low viral load sample containing a drug resistance mutation, the pooling protocol must have equal input from each sample into the pool. Equal amounts of PCR amplicon as measured by spectrophotometry will be used in the pools, allowing us to make the assumption that each subject contributes an equal number of viral particles to the pool. However, the accuracy with which this step can be taken – as well as the accuracy's impact on our proposed method – remain to be studied.

Yet another difficulty that arises in drug resistance surveillance is the systematic, high-magnitude measurement error of the ASPCR assay. As Rowley et al.⁹ show, a sample containing only 2% resistant virus may appear to have as much as 11% resistant virus, whereas a sample containing 10% resistant virus may appear to have as much as 33%, depending on non-drug-resistance-conferring polymorphisms in the primer binding site. Our model accounts for both the bias and for the added uncertainty due to measurement error.

In order to reduce the complexity of this analysis and to make 'drug resistance prevalence' a well-defined parameter, we make an important simplifying assumption and focus on the prevalence of a single point mutation, the nucleotide substitution at position 103 of reverse transcriptase that converts a lysine (nt_{AAA}) to arginine (nt_{AAC}), commonly known as a 'K103N mutation'. We choose to focus on the K103N mutation for a number of reasons. Firstly, K103N confers resistance to NVP and EFV, two important first-line drugs. Secondly, K103N is commonly occurring and can persist for years in the absence of ART drug pressure in individuals with transmitted resistance²⁰. Lastly, Rowley et al.⁹ use the K103N mutation as an example in their investigation of the measurement error associated with the ASPCR assay, and we thus have data with which to calibrate our measurement error model. Throughout, when we refer to 'drug-resistance', we mean the presence of the AAC mutation at location 103. 'Drug resistance prevalence', by extension, will be the population proportion of individuals with one or more copies of this K103N mutation. A practical advantage of focusing on K103N is that it allows us to use ASPCR, an assay that only provides information about a single allele and is thus less expensive than consensus sequencing. We discuss the relaxation of this simplifying assumption in Section 5.

In Section 2 we provide the statistical details of our proposed method. We then provide an in-depth analysis of the likelihood in Section 3, investigating the tradeoffs entailed in choosing a pool size and the problems that may arise if pools are too large. Lastly, in Section 4, we present a simulation study comparing the results generated from our novel method to those obtained from an existing non-model-based estimation method.

2 A Bayesian model for pooled data

In this section, we present a Bayesian model that uses pooled-testing data to estimate the population-level prevalence of drug resistance and its uncertainty.

Let Y_i denote the proportion of resistant virus in the i^{th} pool member's plasma, with $Y_i = 0$ if subject i is not resistant. For subjects who are resistant, we assume that $P(Y > 0)$ is an unknown distribution with mean μ . We denote this distribution P_{y^+} ; we will describe its functional form in Section 2.1.

Let p denote the population proportion of individuals with drug-resistant virus, our parameter of interest:

$$p = P(Y > 0).$$

Then Y 's distribution is a mixture of P_{y^+} and a point mass at zero (see Figure 1):

$$P(y) = pP_{y^+} + (1 - p)I\{y=0\}.$$

Now consider pooling n independent samples of HIV-infected sera. Note that – since the samples have been standardized by viral load – Z is simply the average of the pool members' resistance levels:

$$Z = \frac{1}{n} \sum_{i=1}^n Y_i.$$

The number of resistant individuals in the pool, k , is binomially distributed:

$$k = \sum_{i=1}^n I\{y_i > 0\} \sim \text{Binom}(n, p).$$

The likelihood, $P(Z|k)$, is $1/n$ times a k -fold convolution of random variables with unknown distribution P_{y^+} . The quintessential statistical challenge of pooled analyses is that the number of resistant pool members (k) is not observable, potentially impeding inference on p , our parameter of interest.

In the simplest case of pooled analysis, we would have (a) perfect knowledge of P_{y^+} , and (b) an assay that measures resistance with perfect accuracy. In this case, $E[Z] = E[Y] = p E[P_{y^+}] = p\mu$, which implies that an unbiased estimator of p is given by $\hat{p} = Z/\mu$. The focus of this paper is to relax these two restrictions, allowing for uncertainty about P_{y^+} and for measurement error. We allow for uncertainty about P_{y^+} through its prior, as we describe in Section 2.1, and we introduce a Bayesian measurement error model in Section 2.2 in which the z 's are no longer measured with perfect accuracy and thus become latent variables in our analysis. Although previous work in this area has accounted for measurement error^{15;16;17;18}, knowledge of Y 's distribution has always been assumed^{17;18}.

2.1 Accounting for uncertainty about the distribution of resistance levels

In order to account for uncertainty about P_{y^+} , we embed the likelihood described above in a Bayesian model. An advantage of the Bayesian approach is that it provides a natural framework for allowing this uncertainty to propagate through the model to inference on p , our parameter of interest.

Importantly, note that in the absence of prior information about P_{y^+} , p may not be identifiable. Consider, for example, an observed resistance level of $z = 0.2$ in a pool of $n = 5$ samples. This could represent 5 samples each with a resistance level of $y = 0.2$, which would imply that p was large. However, z could equally well represent one completely resistant sample ($y = 1$) and four non-resistant ($y = 0$) samples, which would imply that p was small (see Figure 2).

We return to this identifiability issue in Section 3. In this toy example, the only way to disentangle these possibilities would be to use prior information about P_{y^+} . Since knowledge of P_{y^+} is critical and will not always be available from the data, an informative prior for $\{Y|Y > 0\}$ is needed. We therefore choose the shape of P_{y^+} using prior knowledge of HIV viral dynamics.

It has been shown that approximately 78% of HIV transmission occurs as a result of infection with a single viral lineage from which all subsequent virus stems²¹. Primary resistance, then, often occurs as the result of infection with a single *resistant* viral lineage. We can thus infer that in the very early stages of infection with transmitted resistance, an individual's entire viral population will be resistant.

Since our goal is to estimate the prevalence of primary resistance among recently infected, treatment-naïve individuals, this fact has important implications for P_{y^+} , the distribution of resistance levels among those individuals with primary resistance. In particular, we propose to restrict the family of priors for $\{Y|Y > 0\}$ to a continuum of truncated normal distributions that are all increasing in density from 0% to 100% resistance.:

$$P_{y^+} = \mathcal{N}(1, \sigma^2) I\{0 \leq y \leq 1\}.$$

As shown in Figure 3, this choice for P_{y^+} reflects the virology of HIV transmission because most individuals enter into this population with 100% resistance and because the population is limited to those who were infected recently (i.e., their virus has not yet had time to revert to wild type²⁰).

This constraint on the prior space helps to solve the identifiability problem depicted in Figure 2. In particular, the large- p -low- μ scenario shown in the top row would be unlikely since the family of $\mathcal{N}(1, \sigma^2) I\{0 \leq y \leq 1\}$ distributions has values of μ in the 0.5 – 1.0 range. We note that in some cases the data may also provide information about σ ; however, as we show in Section 3, this information will be lost if pool sizes are too large. Measurement error also confounds data-driven inference on σ (or equivalently, after a simple

reparameterization, on $\mu = 1 - \frac{\phi(0) - \phi(1/\sigma)}{\Phi(1/\sigma) - \Phi(0)} \sigma$).

We put noninformative flat hyperpriors on p and on σ^{22} .

2.2 Accounting for measurement error

An important generalization allows for measurement error of Z . It is known that standard genotypic assays have a lower bound of detection and that the $_{ASPCR}$ may yield biased results⁹. Thus, rather than observing Z itself, we observe $W = f(Z)$. We need to investigate the properties of $f(\cdot)$ in order to calibrate this relationship and make valid inference on p .

We have gathered calibration data to describe the measurement error that occurs when using $_{ASPCR}$ to quantify levels of the $_{K103N}$ mutation in pooled samples. Samples with known true proportions of resistant virus (0%, 1%, 2.5%, 5%, 7.5%, 10%, and 20%) were assayed via $_{ASPCR}$ in order to obtain estimates of these true values. In this experiment, the $_{K103N}$ mutation was identified against a background of six non-resistant strains. We note, however, that the number and genotype of the pool's non-resistant background members is not expected to impact measurement error. In Figure 4, we display the results, superimposing the following model:

$$\text{logit}(w) \sim \begin{cases} \mathcal{N}(a_0, v_0) & \text{if } z=0 \\ \mathcal{N}(a_1 + bz, v_1) & \text{if } z>0 \end{cases} \quad (1)$$

This model was chosen empirically as it appears to fit the data well. In other settings, alternative models may be preferred. Under measurement error, W becomes the observable data, Z becomes a latent variable, and Equation 1 becomes the likelihood.

2.3 Fitting the full model

Recall that Z 's distribution is $1/n$ times a k -fold convolution of truncated normal random variables with distribution $P_{y^+} = \mathcal{N}(1, \sigma^2) \mathbb{I}\{0 \leq y \leq 1\}$. Note that σ and k thus fully determine the prior for Z , as shown in Figure 5.

We let $\gamma = \{a_0, a_1, v_0, v_1, b\}$ denote the set of measurement error parameters. A complication that arises in implementing the measurement error model is that only the calibration data, and not the observed w 's from the surveillance study, contain information about γ . Best²³ has shown that putting both types of data into a common model may lead to undesirable feedback from the w 's to inference on γ . We thus wish to implement a model in which γ 's posterior distribution is informed by the calibration data alone.

An ad hoc option would be to obtain point estimates of the measurement error parameters using the calibration data, and then to treat these parameter estimates as known constants in subsequent analysis. However, this method ignores uncertainty about γ and may thus yield overly precise estimates of p .

An appealing alternative accounts for this uncertainty using a single probability model for the complete dataset, "cutting feedback"^{24,25} from the observed w 's to γ . This approach artificially drops the terms involving w 's from the conditional distribution of the γ parameters. Although the resulting model is not a full probability model, Best²³ shows that

the resulting inference may nonetheless be preferred. We adopt this strategy in fitting our model.

In order to fit the full model depicted in Figure 5, allowing for measurement error, but cutting feedback from the simulated w 's to the parameters of the measurement error model, we use two separate MCMC S, as we describe in the Appendix. Rather than integrating k out to express Z 's distribution as a function of p , we include both z and k as auxiliary variables that simplify the MCMC . We implement an adaptive Metropolis algorithm that yields good mixing.

3 An analysis of the likelihood with implications for choosing a pool size

In this Section, we describe the likelihood surface under two sets of 'true' parameter values given various pooled and non-pooled designs, in the absence of measurement error. (Note that since measurement error is not being considered in this Section, the term 'likelihood' refers to $P(z)$ rather than $P(w)$ throughout.) We show that pooling can yield more efficient estimates of our parameter of interest, p . We also investigate the nonidentifiability issue introduced in Section 2.1. These results provide insight into the tradeoffs entailed in choosing a pool size and the problems that may arise if pools are too big.

In choosing the true parameter values, we are mindful of the context of HIV drug resistance surveillance in Africa. We set $p = 0.05$ because we intend our approach to serve as a sentinel surveillance method, detecting low levels of drug resistance in new African ART programs, as described in the Introduction. We set μ , the mean of the truncated normal, equal to 0.9 since we expect μ to be near 1 in our population of recently-infected individuals with primary resistance, as described in Section 2.1. These data-generating truths are shown as green dots in Figure 6. As an aside, recall that large μ corresponds to small σ – we parameterize in terms of μ here since it has a more intuitive interpretation.

For each candidate design, rather than simulating datasets given these parameter values, we deterministically set the proportion of resistant individuals equal to p , and we draw values of y evenly from the quantiles of the truncated normal distribution corresponding to $\mu = 0.9$. In this way, by using the data that are most likely given the 'true' parameter values, we prevent sampling variability from clouding our assessment. This allows optimal comparison of likelihood surfaces across designs. We note, however, that because sampling variability is removed and because measurement error is ignored, this analysis may make estimation of p appear deceptively clear-cut. We reintroduce these two complications in the next section.

As depicted in Figure 6, we pool according to 16 candidate designs, creating $N = 10, 50, 100$, or 200 pools of size $n = 1, 10, 20$, or 50. For each of these 16 simulated datasets, we calculate the likelihood on a grid of values such that the x -axis in Figure 6 shows values of p and the y -axis shows values of μ . For each design, the maximum likelihood value ($p, \hat{\mu}$) is shown as a blue dot.

Estimation of p and μ is successful: we do not observe any systematic bias, and all confidence intervals (calculated using an asymptotic variance approximation from Tu et al.¹⁵) cover their respective true values. The only exception is the $\{n = 1, N = 10\}$ case in which we observe zero resistant individuals and are thus wholly unable to identify μ .

For any number N of tests performed, precision increases with the pool size n . These results suggest that we can reduce the cost of surveillance by pooling, and that for a true p of 0.05 – at least in the absence of measurement error – higher degrees of pooling are not detrimental.

Recall from Section 2.1, however, that in certain instances, pooling may yield p nonidentifiable. We now give an example of this potential lack of identifiability and describe the sources of information that our model uses to solve this problem. We repeat the analysis depicted in Figure 6, increasing the true value of p to 0.1 and decreasing the true value of μ to 0.75. The results are shown in Figure 7.

Note that as pool size increases, a ridge develops in the likelihood corresponding to $E[z] = p\mu$. For example, consider the $N = 50$ column: we obtain a precise estimate of p with pool sizes of $n = 10$ and $n = 20$, but for $n = 50$, a ridge-shaped likelihood is associated with decreased precision. In Figure 8 we take a closer look at these three candidate designs. In the first row, we consider the $\{n = 10, N = 50\}$ design. The likelihood is able to distinguish between the data-generating truth of $\{p = 0.1, \mu = 0.75\}$ shown on the left and another point with $p\mu \approx 0.075$ shown on the right because the former has a larger point mass at zero and is ‘bumpier’ in the $z > 0$ range (e.g., values of 0.11 are highly unlikely). These same differentiating characteristics persist in the $n = 20$ case in the second row of the figure. In the third row, however, we see that z ’s likelihood is nearly identical for these two points with a common value of $p\mu$. This is consistent with the ridge observed in the $\{n = 50, N = 50\}$ cell of Figure 7, suggesting that for a true p of 0.1, pools of $n = 50$ are too large.

We conclude that our model uses the $z = 0$ pools to obtain information about p that is not confounded by our uncertainty about μ . We thus benefit statistically from investigating a low-prevalence condition. Given p , our model is able to use the ‘bumpiness’ of the likelihood in the $z > 0$ range to obtain information about μ . We thus benefit from investigating recently-infected individuals who can be expected to have a large value of μ and thus a ‘bumpy’ likelihood. Importantly, arbitrarily large pool sizes are not advisable, since – as the Central Limit Theorem shows and as we see in Figure 8 – high pool sizes will likely lead to mostly non-zero pools of increasingly homogeneous magnitude. Furthermore – consistent with our comparison of Figures 6 and 7 – as p increases over time, smaller pool sizes are needed to ensure identifiability of p .

4 A simulation study: comparison with an existing method

Recall that the previous Section did not account for measurement error nor sampling variability. In this Section, we expand those preliminary findings to the more realistic scenario that includes these complications. We compare the relative strengths of our model-based approach to an existing non-model-based estimator, and we provide guidelines for choosing a pool size.

4.1 Simulation setup

For the purposes of this simulation, we focus attention on our motivating example of sentinel resistance surveillance in Botswana. We thus investigate model performance for very low values of p (0.02, 0.04, and 0.06), consistent with the finding of Bennett et al.⁵ that the

prevalence of transmitted drug resistance had not yet risen above 5% in any of nine completed surveys in Africa. As before, we consider $\mu = 0.75$ and $\mu = 0.90$ to reflect the fact the viral populations of treatment-naïve, recently-infected individuals with primary resistance have not yet had time to revert to wild type. Furthermore, we restrict our analysis to designs that use $N = 50$ pools, consistent with the WHO surveillance protocol.

We consider two candidate ‘true’ data-generating sets of measurement error parameters γ . The first, $(a_0, a_1, b, v_0, v_1) = (-7.7, -3.3, 13.7, 0.6, 0.7)$, estimated from the calibration data depicted in Figure 4, is characterized by minimal overlap between the observed values of w when the latent z is equal to zero vs. when z is positive, implying that measurement error will not confound those pools with $k = 0$ and those pools with $k > 0$. In the second scenario, we introduce additional overlap by subtracting two from the data-generating value of a_1 : $(a_0, a_1, b, v_0, v_1) = (-7.7, -5.3, 13.7, 0.6, 0.7)$, making the $k = 0$ pools harder to distinguish from the $k > 0$ pools.

We simulate data given known ‘true’ parameter values under a variety of candidate pool sizes for each triplet of p , μ , and γ , generating 50 simulated datasets for each parameters-by-pool-size pair of interest. We estimate p separately for each dataset to obtain draws from the sampling distribution of the posterior mean of p . We start the MCMC chains at the parameters’ true values. After a burn-in of 1000 iterations, we run each chain for 2500 more iterations.

4.2 Alternative estimator

For comparison, for each simulated dataset, we also estimate p using an existing non-model-based approach Tu et al.^{15;26}. We’ll refer to this method as the ‘Tu’ method for the sake of brevity. Consider a dichotomous test with sensitivity Se and specificity Sp that simply counts the number of pools m with w greater than some cutoff. The Tu method estimates:

$$\hat{p}_{Tu} = \begin{cases} 0 & \text{if } m/N < 1 - Sp \\ 1 - ((Se - m/N)/(Se + Sp - 1))^{1/n} - \hat{\delta} & \text{if } 1 - Sp \leq m/N \leq Se \\ 1 & \text{if } m/N > Se \end{cases}$$

where $\hat{\delta}$ is an estimate of the bias of the unadjusted estimator, detailed in Tu et al.²⁶. Two appealing features of this method are (a) its ease of implementation and (b) the fact that it does not rely on the truncated normal assumption made by the model described above. The Tu estimator has two potential disadvantages though. First, some information is necessarily lost by dichotomizing the continuous value w . Second, in those cases when $\hat{p}_{Tu} = 0$ or $\hat{p}_{Tu} = 1$, we obtain little information about the true prevalence²⁶. As we’ll see below, the latter case is especially misleading in our setting, since the true value of p is small.

This method is typically implemented assuming that Se , Sp , and the cutoff are known, fixed properties of the assay. However, we wish to reflect the uncertainty in these values due to our imperfect understanding of the measurement error mechanism, as quantified in our uncertainty about the true value of γ . We capture this uncertainty by calculating a draw of $\hat{p}_{Tu}$ for each posterior draw of γ . Let $\gamma^t = \{a_0^t, a_1^t, v_0^t, v_1^t, b^t\}$ be the t^{th} posterior draw of γ , and let $Z_0 \sim \mathcal{N}(a_0^t, v_0^t)$ and $Z_1 \sim \mathcal{N}(a_1^t, v_1^t)$. We choose the cutoff for this iteration to be the value

c^t that gives $Se^t = Sp^t = P(Z_0 > c^t) = P(Z_1 < c^t)$ (since we have no a priori preference between false negatives and false positives). This cutoff is shown as a red line in Figure 4. The t^{th} draw of $p_{\text{TU}}^{\hat{}}$ is then calculated as a deterministic function of $(Se^t, Sp^t, \text{ and } c^t)$.

4.3 Implications for choosing a pool size

Figure 9 shows that, using our new method, pool sizes of 20 are optimal for a range of currently plausible values of p and μ , regardless of the severity of measurement error. In contrast to the results presented in Figure 6, we do not obtain more efficient estimators by increasing n from 20 to 50, especially for larger values of p . We posit that this is because measurement error smooths over the bumpiness in the likelihood that we identified as beneficial in our discussion of Figure 8.

4.4 Comparison of estimators

In Figure 9, we also compare the performance of the two methods. Most strikingly, these results show that the Tu method is not well-suited to the combination of large pool sizes and large values of p nor to the combination of small pool sizes and more severe measurement error. Either of these combinations may yield $p_{\text{TU}}^{\hat{}} = 1$, which in turns yields very high mean square error (MSE). Given a pool size of 1, we may obtain $p_{\text{TU}}^{\hat{}} = 0$. This does not have as severe an impact on MSE , though, because an estimate of zero is not far from the truth in our setting.

We note that, with the exception of those cases contaminated by $p_{\text{TU}}^{\hat{}} = 0$ or 1, the Tu method and our model-based method yield roughly comparable results. To assess potential differences in light of Monte Carlo simulation error, we obtain 95% intervals around the ratio of the MSE s of the two estimators for each combination of p , μ , n , and the measurement error parameters. We use the delta method to approximate the simulation variance of this ratio, accounting for the ‘matched’ nature of the estimates due to the fact that each simulated dataset was analyzed twice, once by each estimator. With only a few exceptions, the small observed differences between the estimators’ performance can be attributed to Monte Carlo variability in our simulation study, implying that the two methods indeed yield comparable results in these cases.

This is in contrast to the scenario when no measurement error is present, in which case our model outperforms the Tu estimator for all pool sizes by leveraging the likelihood’s bumpiness to differentiate among pools with $k = 1$ vs. 2 vs. 3 resistant individuals and thus improve inference on p (results not shown). Under the measurement error models considered here though, we lose this advantage – the bumpiness is smoothed over by the large magnitude of v_1 relative to b , resulting in overlapping observed w values corresponding to $k = 1$ vs. 2 vs. 3, and implying that not much information is lost by dichotomizing w . Since the difference between a_1 and a_0 is large relative to the v ’s, the observed values of w values corresponding to $k = 0$ vs. $k = 1$ do not overlap very much. As a result, measurement error has little impact on inference under the Tu method.

Credible intervals around $p^{\hat{}}$ cover the true value of p 93% of the time under our model-based method, and 85% of the time under the Tu method. The difference between these two values is driven primarily by those cases when the Tu method estimates $p_{\text{TU}}^{\hat{}} = 1$. Note that when the

Tu method estimates $p_{Tu}^{\hat{}} = 0$, the credible interval – calculated as suggested by Louis²⁷ – often covers the truth.

When measurement error is mild, as is the case for the *ASPCR* assay, and with a p as high as 0.06 and a pool size of 20, the probability of obtaining $p_{Tu}^{\hat{}} = 1$ is negligible, so performance of the Tu method will be comparable to our method's. Since it is easy to implement and does not require any assumption about the shape of Y 's distribution, the Tu method may thus be preferred. Over time though, p is expected to increase in Africa as it has in the United States. As this occurs, it will be necessary either to decrease pool sizes or to begin using our model-based estimation approach. Since decreased pool sizes would be associated with increased MSE , our method may be preferred. To make things concrete, note that for a true p of 15% and a pool size of 20, there is a 12% chance of obtaining an undesirable $p_{Tu}^{\hat{}} = 1$ result, even given the mild measurement error that characterizes the *ASPCR* assay. By contrast, our method brings prior information about μ to bear and is thus reliably able to produce informative estimates in this scenario. Furthermore, when surveillance is carried out in a new setting in which prior information about p is not available, our method may be preferred since it is robust to the possibility that p may be large.

As an aside, note that although estimation of p is our primary scientific objective, an important secondary objective is estimation of the prevalence of dominant quasispecies transmitted drug resistance, the proportion of individuals with at least 20% resistant virus:

$p_d = P(Y > 0.2) = p \int_{0.2}^1 P_y + dy$. Although p is the clinically meaningful parameter – since even ‘minor variants’ may be associated with ART failure²⁸ – estimation of p_d is important for public health program comparisons. In particular, standard HIV genotypic resistance assays are only able to detect resistance mutations that are present above the 20% level in a given sample²⁸. For this reason, most published estimates of drug resistance prevalence are in fact estimates of p_d . Therefore, surveillance of p_d will be important for comparability with existing data. Another advantage of our proposed model over the Tu model is that it yields valid estimates of both p and p_d – no additional data need to be collected. Furthermore, MCMC is inherently suited for determining the uncertainty around complex functionals of parameters, such as p_d .

5 Discussion

An inexpensive *ASPCR* assay enables us to decrease the cost of surveillance by (a) focusing resources on a single important point mutation and (b) pooling multiple samples together to obtain more information from each assay performed. Furthermore, a novel model-based approach allows us to bring virologic prior information to bear, helping to render our parameter of interest identifiable in instances when an existing non-model-based estimator fails. This novel-based approach also allows estimation of the prevalence of dominant quasispecies resistance (p_d) at no extra cost.

The methods discussed in this paper provide an important statistical component of a proposed HIV drug resistance surveillance program for Botswana. Our end goal is to create a ‘black box’ statistical software that can be used in the field to obtain precise, low-cost estimates of drug resistance prevalence. In this analysis, we have implemented an adaptive

Metropolis algorithm that converges reliably and mixes well without requiring user intervention²⁹. As always, though, MCMC takes time and may fail to converge, so importance sampling on a bounded space of plausible parameter values may be considered as an alternative.

A potential extension of the methods presented here would generalize them for the multi-dimensional setting. In particular, for the Botswanan context, we may wish to estimate the joint distribution of the six mutations that are most likely to be circulating given the regimens that have been most commonly used in Botswana ($\kappa 103_N$, $\gamma 181_C$, $M 184_V$, $T 215_Y$, $\kappa 70_R$, and $\kappa 65_R$). Monitoring their joint distribution in the population over time would thus be useful for informing optimal first- and second-line standardized regimens for Botswana and for learning about HIV drug resistance epidemiology. We note, however, that this would require six separate ASPCR assays to be performed on each pool, cutting into the cost savings that are such an appealing feature of this method.

It may also be of interest to relax the parametric assumption that $\{Y|Y>0\}$ comes from a truncated normal distribution. In particular, if more flexibility were desired, a hyper prior over distributions could be used in order to allow the data to choose the functional form of this distribution. As shown in Figure 8, however, information on the shape of P_{y^+} may not be available from pooled data if pool sizes are large. More generally, measurement error clouds information on P_{y^+} , making in-depth inference on the shape of the distribution difficult to implement. Furthermore, recall that since primary resistance often occurs as the result of infection with a single resistant viral lineage, we wish to restrict the prior space to a set of distributions on $(0,1)$ that are monotonically increasing. The range of shapes of such distributions is actually quite narrow, with our assumed family of truncated normals covering the space reasonably well. We thus expect that our inference is robust to the this parametric assumption. These methods may be more broadly applicable, possibly allowing for efficiency gains in prevalence estimation of drug-resistant tuberculosis, malaria, influenza, or hepatitis B, for example. We note, however, that the results presented here depend on the mildness of ASPCR measurement error and on the HIV -specific characteristic of infection with a single viral lineage. For these reasons, our recommendations about pool size and about the strengths and weaknesses of our method vs. the Tu method may not apply in other contexts.

Acknowledgments

Funding

The work was supported by the National Institutes of Health [grant numbers R01 AI089350-01, R01 EB006195-19] and the National Science Foundation [grant number DGE 0644491].

References

1. UNAIDS. Report on the global AIDS epidemic. The Joint United Nations Programme on HIV/AIDS. 2010. URL www.unaids.org/globalreport/global_report.htm.
2. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services; 2011. URL www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.

3. Oette M, Kaiser R, Däumer M, Petch R, Fätkenheuer G, Carls H, Rockstroh J, Schmalöer D, Stechel J, Feldt T, et al. Primary HIV drug resistance and efficacy of first-line antiretroviral therapy guided by resistance testing. *Journal of Acquired Immune Deficiency Syndromes*. 2006; 41(5):573–581. [PubMed: 16652031]
4. Myatt M, Bennett D. A novel sequential sampling technique for the surveillance of transmitted HIV drug resistance by cross-sectional survey for use in low resource settings. *Antiviral Therapy, Supplement 2*. 2008; 13:37–48.
5. Bennett D, Myatt M, Bertagnolio S, Sutherland D, Gilks C. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antiviral Therapy, Supplement 2*. 2008; 13:25–36.
6. Kamoto K, Aberle-Grasse J. Surveillance of transmitted HIV drug resistance with the world health organization threshold survey method in Lilongwe, Malawi. *Antiviral Therapy, Supplement 2*. 2008; 13:83–87.
7. Aghokeng A, Vergne L, Mpoudi-Ngole E, Mbangue M, Deoudje N, Mokondji E, Nambei W, Peyou-Ndi M, Moka J, Delaporte E, et al. Evaluation of transmitted hiv drug resistance among recently-infected antenatal clinic attendees in four central african countries. *Antivir Ther*. 2009; 14(3):401–411. [PubMed: 19474474]
8. Vardavas R, Blower S. The emergence of HIV transmitted resistance in Botswana: When will the WHO detection threshold be exceeded? *PLoS ONE*. 2007; 2(1):e152. [PubMed: 17225857]
9. Rowley C, Boutwell C, Lockman S, Essex M. Improvement in allele-specific PCR assay with the use of polymorphism-specific primers for the analysis of minor variant drug resistance in HIV-1 subtype C. *Journal of Virological Methods*. 2008; 149(1):69–75. [PubMed: 18295909]
10. Dorfman R. The detection of defective members of large populations. *The Annals of Mathematical Statistics*. 1943; 14(4):436–440.
11. Pilcher C, Fiscus S, Nguyen T, Foust E, Wolf L, Williams D, Ashby R, O'Dowd J, McPherson J, Stalzer B, et al. Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*. 2005; 352(18):1873–1883. [PubMed: 15872202]
12. Westreich D, Hudgens M, Fiscus S, Pilcher C. Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests. *Journal of Clinical Microbiology*. 2008; 46(5):1785–1792. [PubMed: 18353930]
13. Smith D, May S, Perez-Santiago J, Strain M, Ignacio C, Haubrich R, Richman D, Benson C, Little S. The use of pooled viral load testing to identify antiretroviral treatment failure. *AIDS*. 2009; 23(16):2151. [PubMed: 19730348]
14. Preiser W, Potschka S, Lundershausen A, Haubrich R, Smith D, Van Zyl G. Pooling strategies to reduce the cost of hiv-1 rna load monitoring in a resource-limited setting. *Clinical Infectious Diseases*. 2011; 52(2):264–270. [PubMed: 21288854]
15. Tu X, Litvak E, Pagano M. Screening tests: Can we get more by doing less? *Statistics in Medicine*. 1994; 13:1905–1919. [PubMed: 7846399]
16. Wein L, Zenios S. Pooled testing for hiv screening: Capturing the dilution effect. *Operations Research*. 1996; 44(4):543–569.
17. Zenios S, Wein L. Pooled testing for HIV prevalence estimation: exploiting the dilution effect. *Statistics in Medicine*. 1998; 17(13):1447–1467. [PubMed: 9695191]
18. Graham, D. PhD thesis. Boston, Massachusetts: Harvard University School of Public Health, Department of Biostatistics; 2005. Statistical methods for the analysis of HIV drug-resistance data.
19. Schisterman E, Vexler A, Mumford S, Perkins N. Hybrid pooled–unpooled design for cost-efficient measurement of biomarkers. *Statistics in Medicine*. 2010; 29(5):597–613. [PubMed: 20049693]
20. Little S, Frost S, Wong J, Smith D, Pond S, Ignacio C, Parkin N, Petropoulos C, Richman D. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *Journal of Virology*. 2008; 82(11):5510–5518. [PubMed: 18353964]
21. Abrahams M, Anderson J, Giorgi E, Seoighe C, Mlisana K, Ping L, Athreya G, Treurnicht F, Keele B, Wood N, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *Journal of Virology*. 2009; 83(8):3556–3567. [PubMed: 19193811]

22. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1(3):515–533.
23. Best, N. Technical report. London, UK: Imperial College; 2007. Cutting feedback in Bayesian full probability models. URL www.math.helsinki.fi/openbugs/IceBUGS/Presentations/BestIceBUGS.pdf,
24. Rougier J. Comment on article by Sansó et al. *Bayesian Analysis*. 2008; 3(1):45–56.
25. Lunn D, Best N, Spiegelhalter D, Graham G, Neuenschwander B. Combining MCMC with ‘sequential’ PKPD modelling. *Journal of Pharmacokinetics and Pharmacodynamics*. 2009; 36(1): 19–38. [PubMed: 19132515]
26. Tu X, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV screening. *Biometrika*. 1995; 82(2):287–297.
27. Louis T. Confidence intervals for a binomial parameter after observing no successes. *The American Statistician*. 1981; 35(3):154.
28. Simen B, Simons J, Hullsiek K, Novak R, MacArthur R, Baxter J, Huang C, Lubeski C, Turenchalk G, Braverman M, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes. *The Journal of Infectious Diseases*. 2009; 199(5):693–701. [PubMed: 19210162]
29. Shaby, B.; Wells, M. Exploring an adaptive metropolis algorithm. Tech. rep. 2010.
30. Roberts G, Gelman A, Gilks W. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*. 1997; 7(1):110–120.

Appendix – MCMC algorithm

In order to fit the full model, allowing for measurement error, but cutting feedback from the observed w 's to the parameters of the measurement error model, we use two separate MCMC_S . We begin by ignoring the w 's and estimating the parameters of the measurement error model via MCMC , assigning uniform priors to a_0 , a_1 , b , $\sqrt{v_0}$, and $\sqrt{v_1}$, and using a Gibbs sampler to obtain draws from the posterior distribution of these five parameters.

In each iteration of the second MCMC , we treat one of the γ draws from the first MCMC as a fixed constant. We update the model's remaining parameters – p , σ , k , and z – using an adaptive Metropolis algorithm:

1. Perform a Metropolis update of $\{p, \sigma\}$ by proposing $\{p^*, \sigma^*\}$ from a bivariate normal distribution centered at the previous values of these parameters, with proposal covariance updated every 100 iterations via the adaptive Metropolis algorithm proposed by Shaby and Wells²⁹. If p^* falls outside of $[0,1]$, or if σ^* falls outside of $[0.001, 1]$, reject the proposal. Otherwise, accept the proposal with acceptance probability

$$R = \frac{P(z|\sigma^*, k)P(k|p^*)P(p^*, \sigma^*)}{P(z|\sigma, k)P(k|p)P(p, \sigma)}.$$

A practical difficulty of calculating this acceptance ratio is that the pseudodata likelihood, $P(z|\sigma, k)$, is $1/n$ times a k -fold convolution of truncated normal random variables. For values of $k \geq 2$, this seemingly simple likelihood is not available in closed form. We solve this problem by implementing a novel simulation method that allows us to approximate $P(z|\sigma, k)$ in a computationally-tractable manner.

- For $k = 0$, $P(z|\sigma, k)$ is simply an indicator, $I\{z = 0\}$.

- For $k = 1$, $P(z|\sigma, k)$ is truncated normal: $\mathcal{N}(1/n, \sigma^2/n^2) \mathbb{I}\{0 \leq y \leq 1/n\}$.
- For each possible value of k in $2, \dots, n$, $P(z|\sigma, k)$ is not available in closed form. Rather than resorting to asymptotic approximations of the likelihood (as in Graham¹⁸), we use simulation to estimate $P(z|\sigma, k)$, as depicted in the Figure below and described here:
 - For a given value of σ , simulate k values from the $\mathcal{N}(1, \sigma^2) \mathbb{I}\{0 \leq y \leq 1\}$ distribution.
 - Sum them and divide by n to obtain a single random draw from $P(z|\sigma, k)$.
 - Repeat this process to obtain a histogram of 10,000 draws.
 - Smooth the histogram using a kernel density estimator, setting the density value outside of $[0, k/n]$ to zero.

2. For each pool in turn, perform a Metropolis-Hastings update of $\{k, z\}$:

- a. Propose to decrement or increment k or to keep k constant, with the probabilities of proposing each of these moves updated every 100 iterations during burn-in to obtain theoretically optimal acceptance rates³⁰.

Figure. For $k = 2$, $P(z|\sigma, k)$ is not available in closed form. We approximate the likelihood via simulation by smoothing over a histogram of draws using a kernel density estimator.

- b. Propose z^* from its full conditional distribution given the proposed value of k .
- c. Accept the proposal with the following acceptance probability, using the simulation method described in (1) to calculate $P(z|\sigma, k)$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

$$R=\frac{P(w|z^*,\gamma)P(z^*|\sigma,k^*)P(k^*|p)}{P(w|z,\gamma)P(z|\sigma,k)P(k|p)}.$$

Figure 1.
 $P(y)$ is a mixture distribution with a point mass at zero.

Figure 2.

The value of p may not be identifiable in the absence of prior information.

Figure 3.

Restricting the family of priors for $\{Y|Y > 0\}$ to a continuum of truncated normal distributions that all increase in density from 0% to 100% resistance yields a range of virologically reasonable priors.

Figure 4.

Calibration data illustrate the degree of $_{ASPCR}$ measurement error. We compare the logit of the observed resistance level (w) using $_{ASPCR}$ to the true resistance level in a sample (z). In blue, we show the distribution of points with $z = 0$ ($a_0 = -7.7$, $v_0 = 0.6$); in green we show the linear association between z and logit w for those pools with $z > 0$ ($a_1 = -3.3$, $b = 13.7$, $v_1 = 0.7$). The R^2 of this measurement error model is 87%. The red line depicts a cutoff that we discuss in Section 4.2.

Figure 5.

Information flows through the full model as shown in this directed acyclic graph (DAG). The measurement error model is implemented such that information doesn't feed back from w to the measurement error parameters $\{a_0, a_1, v_0, v_1, b\}$.

Figure 6.

Pooling increases precision. For each of 16 candidate designs corresponding to $N = 10, 50, 100$, or 200 pools of size $n = 1, 10, 20$, or 50, we show the likelihood surface (indicated by red/yellow shading) of a dataset chosen to reflect a true p of 0.05 and a true μ of 0.90. The true parameter values are shown in green; the maximum likelihood estimates (MLEs) are shown in blue. Note that multiple blue dots in the top-left panel indicate multiple MLEs.

Figure 7.

As p increases over time, large pool sizes will no longer provide precise estimates of p . For each of the 16 candidate designs, we calculate the likelihood of a dataset chosen to reflect a true p of 0.10 and a true μ of 0.75. The true parameter values are shown in green; the maximum likelihood estimates are shown in blue.

Figure 8.

We investigate the ridge-shaped likelihood observed in the $\{n = 50, N = 50\}$ cell of Figure 7. As the pool size n gets large, the likelihood curves become indistinguishable for two points on the ridge in the parameter space.

Figure 9.

For each of three prevalences ($p = 0.02, 0.04, 0.06$), two means of Y ($\mu = 0.75, 0.90$), and two degrees of measurement error severity ($a_1 = -5.3, -3.3$), we assess estimator performance as a function of pool size ($n = 1, 10, 20, 50$) for our model and the Tu method. In the left two columns, we show the means and the 2.5th and 97.5th percentiles of the $\hat{p}$ s across simulations. (Note that the y-axis scale differs across rows.) In the right two columns, we compare $\log_{\text{MSE}}$ of $\hat{p}$. Pool sizes of 20 appear optimal in most cases. Differences between the estimators are small, except in those scenarios in which the Tu method estimates p to be 1.