

**Using Interviewer Random Effects to Calculate Unbiased HIV Prevalence Estimates
in the Presence of Non-Response: a Bayesian Approach**

September 2013

Mark E. McGovern¹, Till Bärnighausen², Joshua A. Salomon³ and David Canning⁴

Harvard Center for Population and Development Studies

Abstract

Selection bias in HIV prevalence estimates occurs if refusal to test is correlated with HIV status. Interviewer identity is plausibly correlated with consenting to test, but not with HIV status, allowing a Heckman-type correction that produces consistent HIV prevalence estimates. We introduce an interviewer random effects Bayesian estimator which overcomes the identification problems and corrects for the small sample bias in the existing fixed effects approach. Our new estimator also allows the construction of bootstrapped standard errors and is easy to implement in standard statistical software. Our approach is used to produce new estimates and confidence intervals for HIV prevalence among men in Zambia and Ghana.

Keywords: HIV, Heckman Selection Models, Missing Data, Bayesian Estimation

JEL Codes: C11; I19; J11

This work was supported by The Program on the Global Demography of Aging, which receives funding from the National Institute on Aging, Grant No. 1 P30 AG024409-06.

¹Corresponding Author. Program on the Global Demography of Aging Fellow, Harvard Center for Population and Development Studies, Email: mcgovern@hsph.harvard.edu

²Assistant Professor of Global Health, Department of Global Health and Population, Harvard School of Public Health, Email: tbaernig@hsph.harvard.edu

³Professor of Global Health, Department of Global Health and Population, Harvard School of Public Health, Email: jsalomon@hsph.harvard.edu

⁴Richard Saltonstall Professor of Population Science, Department of Global Health and Population, Harvard School of Public Health, Email: dcanning@hsph.harvard.edu

1. Introduction

Estimates of HIV prevalence in developing countries from serologic testing in nationally representative household surveys have been considered a “gold standard” (Boerma et al., 2003). However, survey-based prevalence estimates have the drawback that consent rates to test for HIV from respondents may be low. For example, in the 2003 Ghana Demographic and Health Survey (DHS) 17% of men declined to give blood for a HIV test, while in Zambia in 2007 the figure was 27%. If the rate of HIV infection is different in the group of individuals who refuse than in those who test, then ignoring this missing data may lead to biased estimates of population prevalence. Existing estimates typically impute data for these individuals under the assumption that non responses are “missing at random”, conditional on the other covariates in the model. This produces population prevalence estimates that are very close to the estimates for the sample that consents to test (Mishra et al., 2008). Single imputation is the method recommended by the World Health Organisation for dealing with missing values. This approach uses the out of sample predicted value (i.e. the predicted value based on observable characteristics of those who have missing data on HIV status) from the independent variables in a model for observed HIV status. This method therefore requires the assumption that there are no unobserved variables that are correlated with both HIV status and consent to test, as do other methods which rely on the missing at random assumption, including propensity score re-weighting and multiple imputation. This assumption is unlikely to hold since a person’s belief about his or her HIV status may be related to actual status and may influence the likelihood of consenting to a test. For example, individuals who already know, or suspect based on evaluation of past behavior, that they are HIV positive may be less likely to consent (Bärnighausen et al., 2012; Reniers and Eaton, 2009; Floyd et al., 2012). If this is the case, there will be a selection bias in population prevalence estimates based on an incorrect assumption of “missing at random”.

An attractive alternative to simple imputation and other single equation methods is to use a selection model that explicitly accounts for the selection process, and procures population estimates that allow for this, such as those proposed by Heckman (1979) and Vella (1998). This approach allows consistent estimation of prevalence, even in the presence of unobservable influences on consent, such as beliefs about HIV status. By accounting for both consent and HIV status equations (individuals first chose whether to test, and it is only conditional on this that we

observe their HIV status) explicitly in the model, the key benefit of the Heckman method is that we do not require the missing at random assumption. Bärnighausen et al. (2011) adopted this approach and found a substantial bias from non-response. In the 2007 Zambia DHS survey, they estimated HIV prevalence among men who refused consent to be 53% using the sample selection model, but only 12% from an imputation model. This provides evidence that existing incidence estimates may be misleading, and that selection on unobservables may be an important source of bias in existing estimates. Hogan et al. (2012) used the same method to produce selection corrected HIV prevalence estimates for 12 African countries, while Clark and Houle (2012a) used the approach to produce corrected estimates for the Agincourt Health and Demographic Surveillance Site in South Africa.

Heckman type selection models require an exclusion restriction for identification, namely a variable which predicts consent but not HIV status. The innovation in Bärnighausen et al. (2011) is to use the identity of interviewers as a plausible variable that is correlated with consenting to test but not with HIV status. Higher quality interviewers may have a personality type (for example the ability to show empathy for the interviewee), or relevant experience, which increases their response rates. This is a testable assumption, and we show in tables 1 and 5 that interviewer identity is highly correlated with consent.

Interviewers are not randomly assigned to participants in DHS surveys, but are assigned to specific regions and tend to be matched to respondents by sex and language. However, as Bärnighausen et al. (2011) argue, sex and language (as well as rich set of other participant characteristics) are recorded in the DHS datasets and can thus be controlled for in the analysis. Once they have been controlled for, the interviewer assignment is plausibly random and uncorrelated with participants' HIV status. Clark and Houle (2012b) confirmed this theoretical claim by undertaking a simulation study of the Bärnighausen et al. (2011) approach in which they set the true prevalence rate in their simulated data and showed that the method provides reasonable estimates in large samples unless the assumption that conditional on observed participant characteristics interviewer identity is uncorrelated with HIV status is violated.

While the underlying model is reasonable, there are a number of potential limitations associated with the Bärnighausen et al. (2011) approach. Firstly, they use interviewer fixed effects as their

key variables that affect consent but not status. These fixed effects, however, are not identified for interviewers who conducted only a very small number of interviews, or who have only successes or failures in consent, or who are the only interviewer to interview people of a particular language in a particular region, making their fixed effects collinear with these characteristics. In these cases, the interviewer fixed effects are pooled to a common value in the Bärnighausen et al. (2011) approach. The assumption of a common value for the interviewer effect on consent in these cases is difficult to justify. In addition, even when interviewer fixed effects are formally identified, this identification may be weak and lead to lack of convergence in model estimates; Clark and Houle (2012b) found that in a large percentage of their simulations the model failed to converge.

Secondly, while maximum likelihood estimators are consistent, they may have a substantial bias in small samples (Green, 2003; Greenland, 2000). In a simulation study Chiburis et al. (2012) demonstrated that the finite sample bias for the maximum likelihood estimate of the bivariate probit model is particularly large when the true probability of the outcome (HIV status in our case) is near a boundary of zero or one. The intuition for this bias is that the maximum likelihood estimator selects the most likely parameter value (the mode of the likelihood distribution) not the expected value of the parameter. When the likelihood function is uni-modal and symmetrical the maximum likelihood estimator is usually unbiased, as for example in the case of estimating regression parameters in the linear model with Gaussian errors where maximum likelihood estimates are identical to the ordinary least squares estimates. However, when the parameter space is bounded, and the true parameter is near the boundary, the likelihood function is usually highly skewed and its mode and mean can be very different.

Finally, the analytical standard errors for the HIV prevalence estimates reported by Bärnighausen et al. (2011) are too small since they do not account for regression parameter uncertainty in constructing the model of consent and HIV status. Bootstrap standard errors have been shown to provide reliable inference in the context of bivariate probit models, and are recommended as an appropriate alternative in this context (Chiburis et al., 2012), with the caveat that in small samples the maximum likelihood estimate of the correlation is often at the boundary of the parameter space – either + 1 or -1 (Butler, 1996) and when this occurs the assumptions required

for asymptotic normality of the maximum likelihood estimator are no longer met and inference is non-standard. In particular, the bootstrap approach fails (Andrews 1999a, Andrews 1999b).

This procedure is difficult to implement with the fixed effects approach, due to the fact that the number of interviewees per interviewer will vary with each bootstrap sample, and a different set of interviewer fixed effects becomes unidentified, requiring rewriting the model with different pooling of interviewer fixed effects for each bootstrap draw. Hogan et al. (2012) addressed this issue by using a parametric simulation approach; however, this approach requires a strong assumption on the correlations between unobserved error terms in the parameterization.

We account for these issues by proposing a random effects approach with Bayesian averaging. The random interviewer effect assumption means that the interviewer “quality” that affects consent to HIV testing, and likely to comprise some of the attributes described above, is taken to be a random draw from a normal distribution instead of a fixed effect. This enables us to estimate the interviewer effects in cases in which it is difficult to identify and estimate fixed effects. As a consequence we avoid arbitrary pooling of interviewer effects and allow straightforward bootstrapping to obtain correct confidence intervals for our HIV prevalence estimates.

The Bayesian averaging approach addresses the bias in the maximum likelihood estimator by including information on all values of the correlation parameter between the error terms in testing and HIV status which have non zero probability. In contrast, the maximum likelihood estimate (by definition the value with the highest posterior probability) ignores less likely values, even where their posterior probability is positive. We use a flat prior on the interval $[-1, 1]$. We then estimate a posterior distribution for the correlation based on the data and our selection model, which allows us to calculate the expected value of the correlation parameter. This approach avoids the problem of boundary estimates, and is feasible to implement even when the likelihood is monotonic.

We illustrate this methodology using data from the Demographic and Health Surveys, and implement our procedure for Zambian and Ghanaian men. We examine Zambian men so that the results of our random effects model can be compared with the fixed effects approach of Bärnighausen et al. (2011); we show the two approaches produce very similar results. We examine Ghanaian men because the low HIV prevalence in Ghana makes estimation potentially

problematic, and as in Hogan et al. (2012), we find a correlation between being HIV positive and testing for this group to be large and positive (close to a correlation of one), so that the maximum likelihood estimate for HIV prevalence among those who refuse to test is zero. Our methodology is designed both to improve upon the existing fixed effects approach and to be easy to implement so that researchers can produce both prevalence estimates and valid confidence intervals given their survey data in a straightforward way.⁵

We have two main empirical results. The first result confirms the finding in Bärnighausen et al. (2011) that the point estimate for the HIV prevalence rate can change a great deal when we allow for response bias. The second result is that, as suggested by Hogan et al. (2012), the analytic standard errors reported by Bärnighausen et al. (2011) are too small. Our bootstrapped confidence intervals for the prevalence rate can be very large in the presence of response bias and uncertainty on the correlation between testing and HIV status. The correlation between agreeing to test and HIV status is fundamentally difficult to estimate, and when we have a large volume of non-response in a survey we are actually quite uncertain of the population prevalence rate. The confidence intervals for the standard imputation model, based on non-response generating data that is missing at random, are very small; this is because the model wrongly assumes that the correlation between testing and status is known with complete certainty to be zero.

The rest of this paper is structured as follows. Section 2 outlines the methodology and theory in more depth. Section 3 describes the data and Section 4 presents the results. Section 5 concludes.

2. The Model

2.1 Bivariate Probit

Following Bärnighausen et al. (2011) we model consent to HIV testing for person i with interviewer j as the observed outcome arising from a latent variable that may interpreted as the propensity to consent to testing:

⁵We provide the STATA computer programs used in our analysis. These can easily be adapted to be used with other surveys.

$$\begin{aligned}
s_{ij}^* &= \beta_s x_{ij} + z_j + u_{ij} \\
s_{ij} &= 1 \text{ if } s_{ij}^* > 0, s_{ij} = 0 \text{ otherwise}
\end{aligned} \tag{1}$$

where s_{ij} is a dummy for agreeing to test, x_{ij} are observed characteristics, z_j are interviewer effects, u_{ij} is random error, and s_{ij}^* is an unobserved latent variable. The equation for HIV status h_{ij} of individual i with interviewer j is:

$$\begin{aligned}
h_{ij}^* &= \beta_h x_{ij} + \varepsilon_{ij} \\
h_{ij} &= 1 \text{ if } h_{ij}^* > 0, h_{ij} = 0 \text{ otherwise} \\
h_{ij} &\text{ observed only if } s_{ij} = 1
\end{aligned} \tag{2}$$

where h_{ij}^* is again a latent variable⁶ and ε_{ij} is an error term. We assume $(u_{ij}, \varepsilon_{ij})$ are bivariate normal, each with mean zero, variance 1, and correlation parameter $\rho = \text{corr}(u_{ij}, \varepsilon_{ij})$.

The key parameter in the model is ρ . If $\rho = 0$ there is no correlation between testing and HIV status. In this case, simple imputation of the HIV status of those who do not test based on their observed characteristics, as in Mishra et al (2008), is possible. If $\rho \neq 0$ however, testing is correlated with HIV status and the predicted probability of being HIV positive of those who refuse to test will be different than the average prevalence rate among people with similar observable characteristics who do test. The key issue in the model is therefore to find robust estimates of ρ .

The inclusion of the interviewer effects, which are assumed to affect consent to testing, but not HIV status, is crucial to the model. Without variables in the selection equation that are excluded from the HIV status equation, the model is identified only by non-linearities, and does not provide robust estimates (Madden, 2008).

Even with a suitable exclusion restriction, there are a number of technical problems associated with estimating the model. The first concerns the estimates of the interviewer effects in z_j . Bärnighausen et al. (2011) estimate each z_j as a parameter on an interviewer dummy, essentially an interviewer fixed effect. Estimation of these interviewer fixed effects may be difficult because

⁶Which can be thought of as reflecting propensity to be infected.

for interviewers who have only successes or failures in obtaining consent, the parameter is not identified. For example, for interviewers who always obtain consent we know their interviewer effect is large and positive, but it can be arbitrarily large as any very large positive interviewer effect above some threshold will predict perfect success, and we cannot distinguish between different effect sizes. In addition, we control for the region and language of the interviewee. However, interviewers usually work in only a small number of regions, and an effort is made to match them with interviewees by language. This may create co-linearity between the interviewer effect and the region and language dummies in the equation, making it difficult or even impossible to estimate the interviewer effects (i.e. weak or no identification of the interviewer parameter). The approach in Bärnighausen et al. (2011) was to limit estimation of individual interviewer effects to those interviewers with more than 50 interviews, and not to use the individual effects of those interviewers who had conducted more than 50 interviews when including them lead to identification problems.⁷ Instead, interviewers with less than 50 interviews or whose inclusion caused identification problems were grouped together as a baseline group with an assumed common interviewer effect.

A conceptual difficulty is that grouping all of the interviewers whose individual interviewer effects may not be identified into a baseline group with the same assumed effect may not be a convincing strategy, because we may be pooling interviewers who always have consent with those who never achieve consent and assuming they have the same average effectiveness in obtaining consent. Thus we do not exploit the fact that the HIV tests made by interviewers who always obtain consent are the most informative since they do not suffer from selection bias. Even if the individual interviewer effect is not fully identified, it may be partially identified (e.g. see Tamer, 2010), so that based on the data we can limit its value to an interval, and the fixed effect approach with pooling for under-identified effects does not exploit this.

A second difficulty is related to drawing bootstrap samples for uncertainty estimation. Bootstrapping is particularly appealing in this setting since we want to know the confidence intervals of the model's predictions as well as for its parameters. Chiburis et al. (2012) have

⁷ The identification problem is usually manifested in practice by the model failing to converge, technically the unidentified parameters have estimated standard errors that are large and increasing at every iteration - the limit is an unbounded standard error on the unidentified parameter.

illustrated that the bootstrap is appropriate in the context of bivariate probit selection models. However, the identification of the interviewer fixed effects depends on the bootstrap sample drawn, and the interviewer fixed effects chosen to ensure identification in the actual sample are often not identified in a bootstrap draw.

2.2 Random Effects

Our approach to this problem is to assume that the unobserved interviewer effects reflect some underlying parametric distribution which describes interviewer quality. If interviewers were assigned randomly to survey households we could simply assume that each interviewer was a random draw from the pool of interviewers. However the systematic matching of interviewers to subjects, particularly by region and language, means that there may be a correlation between interviewer success rate and who they are matched with. Following Mundlak (1978) and Chamberlain (1980), we write the interviewer effect as:

$$z_j = \delta \bar{x}_j + v_j, \bar{x}_j = \sum_i^{n_j} x_{ij} / n_j, v_j \sim N(0, \sigma_z^2) \quad (3)$$

Where \bar{x}_j represents the average characteristics of the n_j people that interviewer j interviews. Excluding these controls could potentially invalidate the exclusion restriction of the model. In order to be consistent, a requirement of random effects models is that there is no correlation between the explanatory variables in the model and the error term (Mundlak, 1978; Chamberlain, 1980). In order for this to hold, we must assume that, controlling for these observable averages, the variation in which across interviewers is caused by their systematic assignment to particular groups, there is no remaining correlation between the error term in equation (3) and the individual level variables in the model. In particular, we assume that interviewers are not systematically assigned to groups of survey participants based on unobservable characteristics but only on the observable characteristics measured by the survey.

The \bar{x}_j are the average characteristics of the set of interviewees a particular interviewer is assigned to. As these depend only on the survey design, they should not enter the HIV status equation⁸.

This assumption of interviewer random effects gives us the selection equation

$$\begin{aligned} s_{ij}^* &= \beta_s x_{ij} + \delta \bar{x}_j + v_j + u_{ij} \\ s_{ij} &= 1 \text{ if } s_{ij}^* > 0, s_{ij} = 0 \text{ otherwise} \end{aligned} \quad (4)$$

In principle we could estimate the system given by (2) and (4) by maximum likelihood. However, this is difficult as we have a selection equation which has a random effect, requiring numerical integration, inside a bivariate probit model.

There is, however, a simple consistent estimator. Dubin and Rivers (1989) show that the bivariate probit model with selection can be estimated by first finding a consistent estimate of the parameters of the selection model, ignoring the covariance of the error terms, and then estimating the parameters of the full model by maximum likelihood, holding the selection equation parameters at their first stage estimates. The procedure is as follows. We firstly estimate the interviewer effects from the selection equation only (stage one), then include these constructed parameters as our exclusion restriction in a Heckman model (stage two). This two-stage approach is consistent, though not fully efficient, and the second stage does not produce the correct analytic standard errors since the interviewer effects are estimated in the first stage but treated as exogenous variables in the second stage. Murphy and Topel (1985) discuss confidence intervals for models using data that are estimated from a first stage. We can implement this approach by first estimating the interviewer effect as shown in equation (3). Equation (4), the selection equation, is then run using the predicted interviewer effect as the exclusion restriction. Assuming random effects avoids the estimation problems of the fixed effects approach. The assumption that the interviewer effects are normally distributed random effects around $\delta \bar{x}_j$ means we have a smaller set of parameters to estimate than in the fixed effects model. A simple method of constructing consistent estimates of the selection equation for the first stage is to estimate a

⁸ If they are included in the HIV status equation they have little effect on our results; they simply reduce the efficiency of estimation.

probit model and compute the predicted random effect \hat{v}_j as the average of the error term for each interviewer.⁹

We use these selection equation estimates to compute $\hat{z}_j = \hat{\delta}\bar{x}_j + \hat{v}_j$ the estimated interviewer effect, where $\hat{\delta}$ is the estimated coefficient on the interviewer average from (3), and \hat{v}_j is the predicted random effect from (3).¹⁰ Having obtained the estimates \hat{z}_j , we can then estimate the full bivariate probit model in (1) and (2) using this estimated interviewer effect as the variable that affects selection but does not appear in the HIV status equation.

Since the first stage is a consistent estimator of the interviewer effect, the two-stage procedure will be consistent. We can address the problem of incorrect standard errors by bootstrapping over the whole two-stage procedure.¹¹ That is to say, for each bootstrap resample we re-estimate the first stage consent, and then run the bivariate probit model with the estimated interviewer effects from the first stage, and calculate the prevalence estimates based on the predicted HIV status of those who refuse to test. For both Zambia and Ghana we compare our estimates of the interviewer effects based on this random effects approach with those found by the fixed effects model.

2.3 Bayesian Averaging

While the random effects approach solves the identification problems in Bärnighausen et al. (2011) and allows bootstrapping, a second problem remains. In small samples, particularly when failure to consent is rare, or the HIV rate is low, it is difficult to estimate the correlation between consent to test and HIV status (Butler, 1996), and the maximum likelihood estimates in the context of a bivariate probit model are biased (Chiburis et al., 2012). In these cases the maximum likelihood estimate of ρ may fail to converge, or produces a result of +1 or -1, on the boundary

⁹ Estimating this first stage as a random effects model is highly computationally intensive and produces almost identical results to the simple probit. The simple probit produces consistent estimates and this is all that is required for the second stage.

¹⁰ Our results are robust to just using \hat{v}_j alone as the interviewer effect rather than the full estimate $\hat{\delta}\bar{x}_j + \hat{v}_j$.

¹¹ An alternative would be to calculate corrected standard errors as in Murphy and Topel (1985).

of the possible parameter space, in which case the assumptions required for asymptotic normality of the maximum likelihood estimator are violated, and standard inference, including bootstrapping, is invalid (Andrews, 1999a; Andrews 1999b). Such a result also has the implication that, in terms of predicted probabilities, everyone who fails to test is either HIV positive with certainty ($\rho = -1$), or HIV negative with certainty ($\rho = +1$), which seems implausible. In these cases other estimators may have better small sample properties than the maximum likelihood estimator (Fosdick and Raftery 2012). In general, maximum likelihood often does not have desirable finite sample properties (Green, 2003; Greenland, 2000) as the estimate is the most likely value (in terms of posterior probability), which gives zero weight to values with lower posterior probabilities, even when those probabilities are positive.

We wish to construct an estimate of ρ that is consistent and also corrects for this small sample bias. The Bayesian approach has particular appeal within a likelihood framework. Take a data set x^* and a parameter vector θ^* . The likelihood of θ^* is simply the probability of the data given these parameters:

$$L(\theta^*) = P(x^* | \theta^*) \quad (5)$$

We are not generally concerned with the likelihood of the observed data given the parameters, we are more interested in the probability of the parameters given the observed data. However, by Bayes rule:

$$P(\theta^* | x^*) = \frac{P(x^* | \theta^*) P(\theta^*)}{P(x^*)} \quad (6)$$

Where $P(\theta^*)$ is our prior on the parameters. If we have a flat prior so $P(\theta^*)$ is the same for every θ^* we have $P(\theta^* | x^*) \propto P(x^* | \theta^*) = L(\theta^*)$, and the maximum likelihood estimate of θ^* is also the estimate that has the highest posterior probability given the data. While this estimate is the most likely parameter value given the data we can construct an alternative Bayesian estimator as

$$\widehat{\theta^*} = E(\theta^* | x^*) = \int \theta^* P(\theta^* | x^*) d\theta^* \quad (7)$$

If the prior is flat we have $P(\theta^*|x^*) \propto L(\theta^*)$, the posterior probability is proportional to the likelihood and we can write our estimator as

$$\widehat{\theta^*} = E(\theta^*|x^*) = \int \theta^* kL(\theta^*) d\theta^* \quad (8)$$

Where k is a normalization factor so that the integral of the likelihood over the parameter space is one and $kL(\theta^*)$ is a probability density function.

This approach essentially takes account of model uncertainty in small samples. The standard maximum likelihood approach chooses the most likely value of θ^* , while our Bayesian approach gives us an average value of θ^* where we average over different models weighted by the probability of the model being correct. This Bayesian averaging gives consistent estimates (the likelihood function asymptotically puts zero weight on incorrect parameters) and is an unbiased estimator by construction under the assumption that the prior is correct. It has good small sample properties in Monte Carlo studies (Fosdick and Raftery 2012). This Bayesian averaging approach may be particularly appropriate where we are interested in predictions, since these predictions take account of uncertainty of the model parameters (Hoeting et al., 1999). This approach is implemented by calculating the likelihood for each value of ρ , and then taking the weighted average of ρ , where the weights are the likelihood values (transformed so that these values integrate to 1).

In principle we could construct Bayesian average estimates for all the parameters in the model. However, in practice the maximum likelihood estimates for most of the parameters in the model are well determined, and it is only the correlation parameter ρ that really poses problems. We

therefore Bayesian average only over ρ , using a concentrated likelihood for each ρ (see appendix for details). Once ρ is estimated by Bayesian averaging we find the maximum likelihood values of the other parameters given this value of ρ . We report both the maximum likelihood estimates and our new Bayesian average estimates together with bootstrapped standard errors and confidence intervals in each case.

As with any Bayesian estimator there is an issue with the choice of a prior distribution. We use a uniform prior for ρ on the interval $[-1,1]$. Given our model, we can estimate the HIV status among those who do not consent to test by calculating a predicted probability of being HIV-positive given the model. We can bootstrap the estimated prevalence rate based on these predictions. This approach corrects both for the model uncertainty (each bootstrap replication estimates a new set of parameter values) as well as for sampling uncertainty since the bootstrap replicates the variation in who is sampled due to the survey design. The DHS surveys are carried out within fixed strata representing urban and rural areas of each region. With each stratum, a cluster of households is randomly selected from a set of possible primary sampling units defined by a preceding census. Our bootstrapping takes account of the stratification and cluster randomization of the survey design by drawing fixed number of clusters (the same as in the original data) from each stratum in each sample.

This approach gives us estimates for those who refuse to consent to test. For individuals who appeared on the household roster and were not contacted or interviewed we follow the standard approach to impute HIV status based on the set of reported characteristics of the person (reported by the household respondent). Bärnighausen et al. (2011) also used a selection model to predict the prevalence rate of this group. In principle rather than separate selection models for contact

and interview, and consenting to test, we should have a sequential model since requesting a test only occurs after people are interviewed (Clark and Houle, 2012a). However both Bärnighausen et al. (2011) and Clark and Houle (2012a) found no evidence of a selection bias among those who were not interviewed, and we simply impute the HIV prevalence for this group based on the assumption they are missing at random and focus on the potential selection bias among those who were interviewed but refused to test.

3. Data

We use data from the nationally representative Demographic and Health Surveys of men aged 15-59. We present results from Zambia (2007) and Ghana (2003). The samples consist of 7,134 men in Zambia, and 5,334 men in Ghana. Table 1 shows the numbers testing for HIV, the numbers who undertook a DHS interview but refused consent to test for HIV, and the number who did not undertake an individual interview and did not test. The majority of people in the latter group are those who could not be contacted and thus could not be offered interview participation and HIV testing. Further details of the sampling design and HIV testing procedure are provided in Bärnighausen et al. (2011). Descriptive statistics for the individual interview samples for Zambia and Ghana are shown in Tables 2 and 3.

Figure 1 shows the distribution of the number of interviews by interviewer for men for Zambia. Approximately 12% (or 750 individuals) had an interviewer with less than 50 interviews in total. This highlights one of the advantages of the random effects approach we adopt in this paper, namely that we are able to estimate an interviewer effect for each of these individuals, even those whose interviewers conducted a single interview. As has been documented previously

(Bärnighausen et al., 2011), there is substantial variation in the effectiveness of interviewers, as measured by the proportion of individuals they succeed in obtaining consent for a HIV test from. Figure 2 shows the distribution of success rate by interviewer for Zambia. The mean success rate for interviewers is 0.8, with a standard deviation of 0.18. Figures 3 and 4 show the distribution of the number of interviews by interviewer and the distribution of interviewer success rate for Ghana (where there is similar variation to Zambia).

4. Results

We begin by estimating the interviewer random effects as the average error term (per interviewer) from a probit model for HIV testing consent where we include a standard set of covariates along with the mean of these variables for each interviewer to capture the effects non-random allocation of interviewers to participants as shown in equation (4). Table 4 presents marginal effects from the probit for HIV consent for Zambia. This table is a single regression for HIV consent where the first column shows the coefficients on the individual level X variables, while the corresponding coefficients on interviewer averages are shown in the second column. At the individual level the following variables affect consent to test: education, location, prior sexually transmitted disease (STDs), age at first intercourse, number of partners, willingness to care for a HIV positive relative, knowing an AIDS victim, and being a smoker. For example, an extra year of education increases the probability of consenting to a HIV test by roughly 0.5 percentage points. Apart from years of education, all interviewer averages are measured in percentages. Many of the interviewer average measures are significant in predicting consent. For example, in Table 4 if the interviewee speaks Lozi the probability of consent to testing increases,

but the effect is not statistically significant. However, the significant coefficient on the average number of interviews with Lozi speakers your interviewer had is positive and statistically significant. This suggests that the Lozi speaking interviewers may have been better than average at obtaining consent. The distribution of fitted predicted random effects, the average of the error terms for each interviewer, from this model are shown in Figure 5.

Table 5 presents marginal effects for the Heckman selection model consent and HIV equations for Zambia. The first two columns give results for the maximum likelihood fixed effects approach as used by Bärnighausen et al. (2011). The middle two columns give results for our maximum likelihood random effects approach, and the final two columns give the results for Bayesian averaging of the random effects model. Controls for age, region, ethnicity, and religion are included but not shown in the table.

The coefficients on the interviewer dummies for each interviewer for the consent equation in the Heckman approach in column 1 are also not shown because of space constraints.¹² In the consent equations for the random effects approach in columns 3 and 5, we report the coefficient on the interviewer effect estimated from the regression shown in Table 4. The first two models in Table 5 are estimated by maximum likelihood. To estimate the third model the Bayesian average we estimate the likelihood on a grid of values of ρ (we use values between -1 and +1 at intervals of 0.01). We then calculate the likelihood of the model and posterior probability of each value of ρ and then find the expected value of ρ based on this probability.¹³ The coefficients reported in table 5 for the Bayesian average are calculated with this value of ρ imposed. Coefficient

¹²See figure 5 in the appendix.

¹³ The posterior probability is calculated by applying a constant to the likelihood for each value of ρ such that these transformed likelihoods integrate to 1.

estimates across the three models are very similar, and in any case our main parameter of interest when estimating correcting prevalence rates is the correlation between the selection and HIV equations.

In Table 6, we present the estimated correlation coefficient ρ for Zambia from the different models. The negative values estimated indicate that those who refuse to test are more likely to be HIV-positive. The maximum likelihood random effects approaches yields a ρ of around -0.50, which is somewhat lower than the 0-.75 obtained from the fixed effects estimator. The Bayesian average estimate is slightly smaller yet at -0.44. A negative value of ρ indicates that male respondents who are HIV positive are less likely to consent to a test in Zambia.

Figure 6 presents the Bayesian averaging posterior for ρ where we use the random effects exclusion restriction and the concentrated likelihood approach described in section 2. Figure 6 also shows maximum likelihood estimate, the Bayesian average estimate and the 95% bootstrap confidence interval for the Bayesian average estimate. The 95% bootstrap confidence interval is calculated using the appropriate centiles from the empirical distribution of the bootstrap estimates. This approach is more appropriate than normal-based approximations when the distribution of the parameter of interest is skewed.

Table 7 presents estimates of HIV prevalence¹⁴ among those who refused to consent to a HIV test, again comparing the fixed effect, random effect and Bayesian average methods. As with the correlation coefficient, the estimates from the random effects model are lower than from the fixed effects model (32% v 52%). The corresponding estimate from the imputation model is 12%. The standard confidence intervals of these prevalence estimates do not take account of the

¹⁴ All of our prevalence estimates are weighted and take account of survey design.

uncertainty in the estimation of interviewer effects, and therefore overstate the precision of the point estimates for ρ (Murphy and Topel, 1985). Chiburis et al. (2012) recommend the use of the bootstrap for inference in the context of selection models to correct for poor coverage of analytic standard errors. In line with these simulation results, we find that the bootstrap confidence interval for the random effects model is almost 10 times as wide as the analytic standard errors for the fixed effects model. Our random effects procedure allows us to implement a bootstrap where we re-estimate the random effects with each iteration to take account of this additional source of parameter uncertainty. We use 1000 iterations to calculate these standard errors, and use the empirical distribution of the replications to calculate the confidence intervals in order to allow for asymmetry. As discussed above, it is not possible to bootstrap standard errors in the fixed effects model due to the difficulties associated with identifying interviewer effects for interviewers with a small number of interviews, which would change with each sample.

Table 8 presents the resulting population estimates. We also compare these results with the estimates from an imputation model, and show the prevalence among those who participated in HIV testing. The imputation model generates an estimate of 12%, which is substantially smaller than the 20% estimate from the fixed effects model. The random effects and Bayesian average models give estimates that fall in between imputation and the fixed effects model (around 16%). Comparing the bootstrap and analytic standard errors again highlights how the precision of these estimates is greatly overstated with standard approaches which do not account for uncertainty in the estimation of ρ .

We next turn to estimating the models for men in Ghana, where the standard maximum likelihood estimator is difficult to implement because of the low HIV prevalence. In addition, the

small sample size may induce bias in conventional maximum likelihood estimates. Table 9 presents marginal effects for the selection model used to calculate the interviewer effects¹⁵ and Table 10 gives the results of the three models for Ghana. As with Zambia, we compare results from the three approaches.¹⁶

Figure 10 demonstrates that for men in Ghana the concentrated likelihood function is approximately monotonic in ρ , resulting in maximum likelihood estimates which are close to the boundary of the parameter space. The maximum likelihood estimate places 0 weight on all the values in the left hand tail of the posterior distribution in figure 10 as they are less probable than the most likely value. To find an unbiased estimate we integrate over the posterior to account for all values of ρ which have positive probability. Under the assumption that the prior is correct, this will result in an unbiased estimate for ρ . We obtain a value of around 0.6, indicating that individuals with HIV are more likely to consent to testing in Ghana, at least for men.¹⁷ Table 11 shows our estimate of ρ our confidence intervals using each model. The positive value of ρ means that those who are HIV positive are more likely to test, and our population prevalence estimates are lower than those obtained from imputation methods. The positive selection we find in Ghana is in contrast to the negative selection we find in Zambia. Hogan et al. (2012) also find positive selection in Ghana and some other countries. Differences in selection mechanisms may reflect differences in HIV prevalence and other country specific factors. For example, being HIV positive itself is unlikely to be a common reason for non-consent in a country with very low HIV

¹⁵Figure 8 shows the distribution of the interviewer random effects for Ghana.

¹⁶Coefficients on the interviewer fixed effects in the relevant consent equation for Ghana are presented in table 10 in the appendix.

¹⁷The fact that the random effects maximum likelihood estimate lies outside the bootstrap confidence interval for the Bayesian average estimate reflects the fact that the posterior distribution has a long left hand tail which is not accounted for by the standard maximum likelihood estimator, and that we use the empirical distribution of the bootstrap estimates to allow for asymmetry when calculating the confidence interval.

prevalence. Tables 12 and 13 report our prevalence estimates for men who refuse to test and for the whole population. Relatively narrow confidence intervals reflect the low HIV prevalence in Ghana and the lack of potential variation in HIV status in replicate samples when the point estimate for the group who refuse consent is close to 0. For the random effects model and the Bayesian averaging approach we find a prevalence rate of approximately 1.4%, compared to 1.6% in the case of the imputation model.

5. Discussion and conclusions

This paper confirms that non-response can be an important source of bias in HIV prevalence estimates taken from household surveys. We innovate by proposing a random effects model which improves on previous fixed effects approaches by allowing us to estimate interviewer effects, even for interviewers with a limited number of interviews. In addition, we propose a Bayesian averaging procedure which facilitates the estimation of the correlation between HIV consent and HIV status, and allows for unbiased estimates in small samples. Using data from the 2007 Zambian Demographic and Health Survey, we find that men with HIV are less likely to consent to a HIV test. For Ghana, we find that conventional methods slightly overstate HIV prevalence.

Perhaps the most important result we find is that the corrected confidence intervals around the HIV prevalence point estimate can be very wide. These wide confidence intervals accurately reflect the fact that it is very difficult to correct statically for the bias that may occur when many people refuse to test. As long as consent rates are as low as they currently tend to be in many nationally representative HIV surveys in developing countries, uncertainty in HIV prevalence

estimation will likely remain high. It is important not to understate this uncertainty and our approach provides the first practical solution to account for both sampling and parameter uncertainty in the estimation of HIV prevalence confidence intervals.

A key assumption in the model is the exclusion restriction, which requires interviewer identity to be uncorrelated with HIV status. While it is plausible that interviewer allocation should only be affected by survey design, this is impossible to prove conclusively. As the resulting HIV prevalence prediction relates to a population for whom we never get to observe true HIV status, it is important to independently validate the model. We are therefore working to obtain objective data in the form of mortality records with which we can do so. This limitation, coupled with the wide confidence intervals we find, points towards an urgent need to improve HIV survey design and execution so as to increase the consent rates to reduce the uncertainty in HIV prevalence estimates that is induced by selection bias.

References

- Andrews, D. W. (1999). "Estimation when a parameter is on a boundary." *Econometrica* 67(6): 1341-1383.
- Andrews, D. W. (2000). "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space." *Econometrica* 68(2): 399-405.
- Bärnighausen, T., Bor, J., Wandira-Kazibwe, S., & Canning, D. (2011). Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. *Epidemiology*, 22(1), 27-35.
- Bärnighausen T, Tanser F, Malaza A, Herbst K, Newell ML (2012). HIV status and participation in HIV surveillance in the era of antiretroviral treatment: a study of linked population-based and clinical data in rural South Africa. *Tropical Medicine & International Health*, 17(8): e103-111.
- Boerma, J. T., Ghys, P. D., & Walker, N. (2003). Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. *The Lancet*, 362(9399), 1929-1931.
- Butler, J. S. (1996). "Estimating the correlation in censored probit models." *The Review of Economics and Statistics*: 356-358.
- Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1), 225-238.
- Chiburis, R. C., J. Das, et al. (2012). A practical comparison of the bivariate probit and linear IV estimators. *Economics Letters* 117(3): 762-766.
- Clark, S.J. & Houle B. (2012a). An Application of the Biprobit Heckman Selection Model to Correct Estimates of HIV Prevalence from Sample Surveys at the Agincourt HDSS in South Africa, Working Paper No. 119, Center for Statistics and the Social Sciences, University of Washington.

Clark, S.J. & Houle B. (2012b). Evaluation of Heckman Selection Model Method for Correcting Estimates of HIV Prevalence from Sample Surveys via Realistic Simulation. Working Paper No. 120, Center for Statistics and the Social Sciences, University of Washington.

Dubin, J. A., & Rivers, D. (1989). Selection Bias in Linear Regression, Logit and Probit Models. *Sociological Methods & Research*, 18(2-3), 360-390. doi: 10.1177/0049124189018002006

Floyd, S., Molesworth, A., Dube, A., Crampin, A. C., Houben, R., Chihana, M., . . . Glynn, J. R. (2012). Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias. *AIDS, Publish Ahead of Print*, 10.1097/QAD.1090b1013e32835848ab.

Fosdick, B. K., & Raftery, A. E. (2012). Estimating the Correlation in Bivariate Normal Data With Known Variances and Small Sample Sizes. *The American Statistician*, 66(1), 34-41.

Greene, W. H. (2003). *Econometric Analysis*, Fifth Edition, Pearson Education India.

Greenland, S. (2000). Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics* 1(1): 113-122.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4), 382-401.

Hogan, Daniel R., Joshua A. Salomon, David Canning, James K. Hammitt, Alan M. Zaslavsky, and Till Bärnighausen, "National HIV prevalence estimates for Sub-Saharan Africa: controlling selection bias with Heckman-type selection models," *Sexually Transmitted Infections*, 2012, Vol. 88(Supp 2), pp i17-i23.

Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics*, 27(2), 300-307. doi: 10.1016/j.jhealeco.2007.07.001

Mishra, V., Barrere, B., Hong, R., & Khan, S. (2008). Evaluation of bias in HIV seroprevalence estimates from national household surveys. *Sex Transm Infect*, 84 Suppl 1, i63-i70. doi: 10.1136/sti.2008.030411

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1), 69-85.

Murphy, K. M., & Topel, R. H. (2002). Estimation and Inference in Two-Step Econometric Models. *Journal of Business & Economic Statistics*, 20(1), 88-97. doi: 10.1198/073500102753410417

Reniers, G., & Eaton, J. (2009). Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS*, 23(5), 621-629. doi: 10.1097/QAD.1090b1013e3283269e3283213.

Tamer, E. (2010). Partial identification in econometrics. *Annu. Rev. Econ.*, 2(1), 167-195.

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources*, 33(1), 127-169.

Tables

Table 1 Sample Size for HIV Estimation

	Zambia	Ghana
Observed HIV Status	5,163	4,271
Missing HIV Status (Consent Refused)	1,318	743
Missing HIV Status (No Contact)	653	320
Total	7,134	5,334

Table 2 Descriptive Statistics (Eligible Men) for Zambia DHS 2007 (Individual Sample)

	No.	%		No.	%
Age Category			Would Respondent Care for HIV Relative?		
15-19	1,410	21.69	No	347	5.34
20-24	1,071	16.48	Yes	6,146	94.66
25-29	988	15.2	<i>Total</i>	6,493	100
30-34	933	14.35			
			Does Respondent Know Anyone Who Died of AIDS?		
35-39	733	11.28	No	2,866	44.19
40-44	471	7.25	Yes	3,619	55.81
45-49	397	6.11	<i>Total</i>	6,485	100
50-54	294	4.52			
55-59	203	3.12			
<i>Total</i>	6,500	100	Ever Had HIV Test		
			No	4,986	76.75
Years of Education			Yes	1,510	23.25
0	309	4.77	<i>Total</i>	6,496	100
1-7	3,118	48.14			
8+	3,050	47.09	Smoker		
<i>Total</i>	6,477	100	No	4,951	76.25
			Yes	1,542	23.75
Wealth Quintile			<i>Total</i>	6,493	100
Poorest	1,145	17.62			
2 nd poorest	963	14.82	Alcohol Drinker		
3 rd poorest	1,315	20.23	No	3,876	59.69
4 th poorest	1,600	24.62	Yes	2,618	40.31
Wealthiest	1,477	22.72	<i>Total</i>	6,494	100
<i>Total</i>	6,500	100			
			Ethnicity		
Location			Bemba	1,117	17.18
Capital, large city	597	9.18	Lunda (Luapula)	152	2.34
Small city	469	7.22	Lala	166	2.55
Town	1,765	27.15	Ushi	149	2.29
Countryside	3,669	56.45	Lamba	121	1.86
<i>Total</i>	6,500	100	Tonga	766	11.78
			Luvale	160	2.46
Region			Lunda (Northwestern)	311	4.78
Central	600	9.23	Mbunda	124	1.91
Copperbelt	812	12.49	Kaonde	249	3.83
Eastern	857	13.18	Lozi	477	7.34
Luapula	560	8.62	Chewa	488	7.51
Lusaka	962	14.8	Nsenga	313	4.82

Northern	715	11	Ngoni	323	4.97
Northwestern	630	9.69	Mambwe	183	2.82
Southern	774	11.91	Namwanga	176	2.71
Western	590	9.08	Tumbuka	295	4.54
<i>Total</i>	6,500	100	Other	930	14.31
			<i>Total</i>	6,500	100
Marital Status			Language		
Never married	2,546	39.17	English	465	7.15
Currently married	3,630	55.85	Bemba	2,374	36.52
Formerly married	324	4.98	Lozi	611	9.4
<i>Total</i>	6,500	100	Nyanja	1,955	30.08
			Tonga	482	7.42
Ever Had STD			Other	613	9.43
No	6,165	95.09	<i>Total</i>	6,500	100
Yes	318	4.91			
<i>Total</i>	6,483	100	Number of Partners		
			None	1,595	24.61
Age at First Intercourse			1	3,921	60.49
Never Had Sex	908	13.98	2+	966	14.9
15 or Younger	2,016	31.04	<i>Total</i>	6,482	100
>15	3,571	54.98			
<i>Total</i>	6,495	100	Religion		
			Catholic	1,347	20.76
Had High Risk Sex			Protestant	4,932	75.99
No	4,707	72.62	Muslim	211	3.25
Yes	1,775	27.38	<i>Total</i>	6,490	100
<i>Total</i>	6,482	100			
			Used Condom Last Intercourse		
No	5,380	82.9			
Yes	1,110	17.1			
<i>Total</i>	6,490	100			

Table 3 Descriptive Statistics (Eligible Men) for Ghana DHS 2003 (Individual Sample)

	No.	%		No.	%
Age Category			Number of Partners		
15-19	1,096	21.85	None	1,707	34.07
20-24	697	13.9	1	2,824	56.37
25-29	715	14.26	2+	479	9.56
30-34	632	12.6	<i>Total</i>	5,010	100
35-39	526	10.49			
			Used Condom Last Intercourse		
40-44	407	8.12	No	4,451	88.86
45-49	435	8.67	Yes	558	11.14
50-54	302	6.02	<i>Total</i>	5,009	100
55-59	205	4.09			
Total	5,015	100	Respondent Would Care for HIV Relative		
			No	1,389	27.71
Years of Education			Yes	3,624	72.29
0	1,178	23.55	<i>Total</i>	5,013	100
1-7	1,110	22.19			
8+	2,714	54.26	Respondent Knows Anyone Who Died of AIDS		
			No	3,052	60.97
<i>Total</i>	5,002	100	Yes	1,954	39.03
Wealth Quintile			<i>Total</i>	5,006	100
Poorest	1,221	24.35	Ever Had HIV Test		
Poorer	953	19	No	4,580	91.34
Middle	883	17.61	Yes	434	8.66
Richer	906	18.07	<i>Total</i>	5,014	100
Richest	1,052	20.98			
<i>Total</i>	5,015	100	Religion		
Region			None	358	7.14
Western	457	9.11	Roman Catholic	794	15.84
Central	300	5.98	Anglican	47	0.94
Greater Accra	621	12.38	Methodist	301	6
Volta	386	7.7	Presbyterian	361	7.2
Eastern	453	9.03	Other Christian	1,785	35.61
Ashanti	785	15.65	Moslem	1,050	20.95
BrongAhafo	593	11.82	Traditional/Spiritualist	317	6.32
Northern	638	12.72	<i>Total</i>	5,013	100
Upper West	387	7.72			
Upper East	395	7.88			
<i>Total</i>	5,015	100			

Marital Status					
Never Married	2,002	39.92	Smoker		
Currently Married	2,726	54.36	No	4,393	87.63
Formerly Married	287	5.72	Yes	620	12.37
<i>Total</i>	5,015	100	<i>Total</i>	5,013	100
Ever Had STD			Ethnicity		
No	4,834	96.8	Akan	2,025	40.38
Yes	160	3.2	Ga/Dangme	338	6.74
<i>Total</i>	4,994	100	Ewe	614	12.24
Age at First Intercourse			Guan	191	3.81
Never Had Sex	1,171	23.38	Mole-Dagbani	1,235	24.63
15 or Younger	511	10.2	Grussi	157	3.13
>15	3,327	66.42	Gruma	188	3.75
<i>Total</i>	5,009	100	Other	267	5.32
Had High Risk Sex			<i>Total</i>	5,015	100
No	3,931	78.46	Language		
Yes	1,079	21.54	English	1,351	26.94
<i>Total</i>	5,010	100	Akan	2,804	55.91
			Ga	53	1.06
			Ewe	270	5.38
			Nzema	34	0.68
			Dagbani	203	4.05
			Other	300	5.98
			<i>Total</i>	5,015	100

Table 4 Marginal Effects from Probit for HIV Test Consent for Zambian Men

Variables	HIV Test Consent	
	Individual Level Variables	Interviewer Level Averages
Years of Education	0.00475** (0.002)	-0.0001 (0.000)
Wealth Category		
Poorest	-0.00731 (0.031)	-0.0086*** (0.002)
Poorer	-0.02721 (0.031)	-0.0046 (0.003)
Middle	-0.03732 (0.027)	-0.0059 (0.004)
Richer	-0.00842 (0.019)	-0.0109*** (0.004)
Location		
Small city	0.12471*** (0.037)	0.0039 (0.004)
Town	0.14458*** (0.044)	0.0036 (0.004)
Countryside	0.20348*** (0.062)	0.0008 (0.003)
Marital Status		
Never Married	0.01165 (0.028)	0.0036 (0.003)
Currently Married	0.03447 (0.031)	-0.0018 (0.004)
Had STD	0.04105* (0.022)	0.0036 (0.003)
Age at First Intercourse		
15 or Younger	0.04986** (0.020)	0.0043 (0.003)
>15	0.03778* (0.022)	0.0013 (0.003)
Had High Risk Sex	0.05241** (0.025)	-0.0038 (0.004)
Number of Partners		
None	0.05095** (0.025)	-0.0078** (0.004)
2+	0.01516 (0.021)	0.0021 (0.003)
Used Condom Last Intercourse	-0.00033 (0.016)	-0.0005 (0.002)
Would Care for HIV Relative	0.05757** (0.024)	0.0030 (0.003)

Knows Someone Who Died of AIDS	0.03795*** (0.011)	-0.0011 (0.001)
Previously HIV Tested	0.00504 (0.013)	-0.0038 (0.002)
Smoker	0.03644*** (0.012)	0.0041** (0.002)
Drinks Alcohol	0.01008 (0.012)	-0.0064*** (0.002)
Language		
English	-0.04601 (0.062)	0.0079 (0.008)
Bemba	-0.00834 (0.061)	0.0037 (0.009)
Lozi	0.04688 (0.056)	0.0130** (0.007)
Nyanja	0.05083 (0.055)	0.0036 (0.009)
Tonga	0.06172 (0.062)	0.0060 (0.008)
Observations	6,416	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Age, region, religion, and ethnicity are included in the regression but are not shown in the table

Note to table 4: Each column is for the same regression where HIV consent is regressed on the X variables shown in the table and the corresponding average (in percent) for each interviewer.

Table 5 Marginal Effects for Heckman Selection Models (Men Zambia 2007)

Variables	Maximum Likelihood Fixed Effects		Maximum Likelihood Random Effects		Bayesian Average Random Effects	
	Consent Equation	HIV Equation	Consent Equation	HIV Equation	Consent Equation	HIV Equation
Interviewer Effect				0.2845*** (0.028)		0.2857*** (0.028)
Years of Education	0.0046** (0.002)	0.0012 (0.002)	0.0046** (0.002)	0.0019 (0.002)	0.0047** (0.002)	0.0020 (0.002)
Wealth Category						
Poorest	-0.0172 (0.020)	0.0486** (0.020)	-0.0199 (0.020)	0.0463** (0.018)	-0.0198 (0.020)	0.0451** (0.018)
Poorer	-0.0262 (0.021)	0.0528*** (0.020)	-0.0299 (0.021)	0.0479*** (0.018)	-0.0298 (0.021)	0.0465*** (0.018)
Middle	0.0028 (0.025)	0.0645*** (0.024)	-0.0017 (0.024)	0.0630*** (0.022)	-0.0017 (0.024)	0.0620*** (0.021)
Richer	0.0055 (0.031)	0.0503* (0.028)	0.0072 (0.030)	0.0493* (0.025)	0.0071 (0.030)	0.0485* (0.025)
Location						
Small City	0.1529** (0.067)	-0.0589 (0.045)	0.1626** (0.064)	-0.0351 (0.035)	0.1622** (0.064)	-0.0308 (0.032)
Town	0.1605*** (0.061)	-0.0764** (0.038)	0.1649*** (0.059)	-0.0495* (0.027)	0.1647*** (0.059)	-0.0449* (0.024)
Countryside	0.1911*** (0.063)	-0.1289*** (0.039)	0.1972*** (0.059)	-0.0956*** (0.028)	0.1967*** (0.059)	-0.0896*** (0.024)
Marital Status						
Currently Married	0.0105 (0.029)	0.0787*** (0.030)	0.0220 (0.028)	0.0791*** (0.027)	0.0221 (0.028)	0.0779*** (0.027)
Formerly Married	-0.0146 (0.028)	0.2125*** (0.026)	-0.0093 (0.028)	0.1992*** (0.025)	-0.0096 (0.028)	0.1950*** (0.023)
Had STD	0.0407 (0.025)	0.1079*** (0.023)	0.0437* (0.025)	0.1073*** (0.019)	0.0437* (0.025)	0.1064*** (0.019)
Age at First Intercourse						
15 or Younger	0.0502** (0.022)	-0.0422 (0.035)	0.0485** (0.021)	-0.0254 (0.031)	0.0487** (0.021)	-0.0228 (0.030)
>15	0.0374* (0.022)	-0.0475 (0.034)	0.0356* (0.021)	-0.0334 (0.030)	0.0357* (0.021)	-0.0312 (0.030)
Had High Risk Sex	0.0531* (0.027)	-0.0275 (0.025)	0.0546** (0.027)	-0.0199 (0.022)	0.0544** (0.027)	-0.0184 (0.021)
Number of Partners						
1	-0.0433 (0.027)	-0.0202 (0.029)	-0.0520* (0.027)	-0.0266 (0.026)	-0.0521* (0.027)	-0.0272 (0.025)
2+	-0.0259 (0.042)	0.0343 (0.039)	-0.0371 (0.042)	0.0293 (0.034)	-0.0371 (0.042)	0.0282 (0.033)
Condom Last Intercourse	-0.0079 (0.016)	0.0629*** (0.015)	-0.0017 (0.015)	0.0586*** (0.014)	-0.0014 (0.015)	0.0575*** (0.013)
Would Care for HIV Relative	0.0487** (0.020)	0.0170 (0.033)	0.0524*** (0.020)	0.0285 (0.030)	0.0525*** (0.020)	0.0298 (0.029)

Know Someone Who Died of AIDS	0.0369*** (0.010)	-0.0144 (0.014)	0.0370*** (0.010)	-0.0081 (0.011)	0.0370*** (0.010)	-0.0070 (0.011)
Previously HIV Tested	0.0053 (0.013)	0.0340*** (0.013)	0.0050 (0.013)	0.0333*** (0.011)	0.0050 (0.013)	0.0328*** (0.011)
Smoker	0.0371*** (0.013)	-0.0262* (0.016)	0.0371*** (0.013)	-0.0197 (0.013)	0.0370*** (0.013)	-0.0183 (0.013)
Drinks Alcohol	0.0097 (0.012)	0.0186 (0.012)	0.0106 (0.012)	0.0184* (0.011)	0.0105 (0.012)	0.0182* (0.011)
Language						
English	0.0276 (0.028)	0.0028 (0.033)	0.0350 (0.029)	0.0075 (0.029)	0.0351 (0.029)	0.0081 (0.029)
Bemba	0.0953 (0.061)	0.0666 (0.049)	0.0916 (0.058)	0.0816* (0.042)	0.0919 (0.058)	0.0826** (0.041)
Lozi	0.0980*** (0.030)	-0.0076 (0.031)	0.0967*** (0.028)	0.0079 (0.026)	0.0965*** (0.028)	0.0099 (0.025)
Nyanja	0.1017** (0.046)	-0.0177 (0.043)	0.1133*** (0.036)	0.0035 (0.035)	0.1132*** (0.036)	0.0066 (0.034)
Tonga	0.0376 (0.048)	-0.0119 (0.064)	0.0453 (0.051)	-0.0132 (0.063)	0.0452 (0.051)	-0.0128 (0.062)
Observations	6,416	6,416	6,416	6,416	6,416	6,416

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note to table 5: Controls for age, region, ethnicity, and religion are included in each regression equation but not shown. The model is a bivariate probit for HIV status and consent to a HIV test. Data are for men who completed an interview. The first model uses interviewer fixed effects as the exclusion restriction. The coefficients from the interviewer effects are not shown in the table (see figure 5). The second model uses the two step random effects procedure where HIV consent is regressed on the X variables, along with the mean values for each interviewer (see table 4). The average error term for each interviewer is added to the predicted value of the interviewer means, and used as the exclusion restriction in the HIV regression. The final model is the Bayesian averaging procedure using the same exclusion restriction. Coefficients are obtained by restricting the value of RHO to its Bayesian estimate and running the bivariate probit. Source: DHS Zambia 2007 (men).

Table 6 Correlation Coefficient Estimates for Men in Zambia

Correlation Coefficient for Refused Consent	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
Fixed Effects Model	-0.7527	-0.9397	-0.2189		
Random Effects Model	-0.5018	-0.7267	-0.1777	-0.6769	0.2282
Bayesian Average Model	-0.4396			-0.6378	0.4319

Note to table 6: The table shows the estimated correlation coefficient between consent and HIV status for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Source: DHS Zambia 2007 (men).

Table 7 HIV Rate among Men who Refused to Test in Zambia

HIV Rate for Refused Consent	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
Fixed Effects Model	0.5195	0.4987	0.5404		
Random Effects Model	0.3203	0.3031	0.3376	0.0627	0.4268
Bayesian Average Model	0.2860	0.2656	0.2977	0.0317	0.3977
Imputation Model	0.1171	0.1078	0.1263		

Note to table 7: The table shows the estimated HIV prevalence rate among individuals who refused consent for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Also shown is the HIV rate using an imputation model for men who refused consent. Source: DHS Zambia 2007 (men).

Table 8 HIV Rate among Men in Zambia

HIV Rate	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
All Men - Fixed Effects Model	0.2013	0.1896	0.2130		
All Men - Random Effects Model	0.1627	0.1526	0.1728	0.1096	0.1842
All Men - Bayesian Average Model	0.1552	0.1454	0.1650	0.1024	0.1789
Men with Non-Missing Data	0.1213	0.1099	0.1327		
Men with No Contact - Imputation Model	0.1525	0.1424	0.1626		
All Men - Imputation Model	0.1233	0.1144	0.1322		

Note to table 8: The table shows the estimated HIV prevalence rate among the population for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Results from an imputation model are also shown, along with estimates only using those without missing data. Source: DHS Zambia 2007 (men).

Table 9 Marginal Effects from Probit for HIV Test Consent for Ghanaian Men

Variables	HIV Test Consent	
	Individual Level Variables	Interviewer Level Averages
Years of Education	0.00224* (0.001)	-0.0002 (0.000)
Wealth Category		
Poorest	0.09640*** (0.016)	0.0136** (0.007)
Poorer	0.08063*** (0.014)	0.0200*** (0.001)
Middle	0.05346*** (0.015)	0.0417*** (0.007)
Richer	0.03687*** (0.014)	0.0376*** (0.006)
Marital Status		
Currently Married	0.02424 (0.022)	-0.0035 (0.004)
Formerly Married	-0.04371 (0.028)	0.0084 (0.011)
Had STD	0.03830* (0.021)	0.0529*** (0.009)
Age at First Intercourse		
Never Had Sex	0.00918 (0.021)	-0.0578*** (0.009)
15 or Younger	-0.02929 (0.018)	-0.0569*** (0.010)
Had High Risk Sex	0.00272 (0.019)	0.0092 (0.007)
Number of Partners		
1	0.00304 (0.022)	-0.0331*** (0.005)
2+	-0.00752 (0.031)	0.0218** (0.010)
Condom Last Intercourse	0.00045 (0.017)	-0.1715*** (0.024)
Would Care for HIV Relative	0.02213* (0.013)	0.0124*** (0.001)
Know Someone Died of AIDS	0.01538 (0.010)	-0.0424*** (0.006)
Previously HIV Tested	0.00564 (0.017)	-0.0097*** (0.003)
Smoker	-0.00928 (0.017)	-0.0233*** (0.007)
Language		
Akan	0.04955***	-0.0147***

	(0.017)	(0.001)
Ga	0.06397**	0.0282***
	(0.026)	(0.005)
Ewe	0.02522	0.0175***
	(0.038)	(0.005)
Nzema	0.04300	0.0112*
	(0.047)	(0.007)
Dagbani	0.01586	-0.0241***
	(0.049)	(0.007)
Other	-0.03509	
	(0.030)	
Observations	4,955	
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		
Note: Age, region, religion, and ethnicity are included in the regression but not shown in the table		

Note to table 9: Each column is for the same regression where HIV consent is regressed on the X variables shown in the table and the corresponding average (in percent) for each interviewer.

Table 10 Marginal Effects for Heckman Selection Models (Ghana 2003)

Variables	Maximum Likelihood Fixed Effects		Maximum Likelihood Random Effects		Bayesian Average Random Effects	
	Consent Equation	HIV Equation	Consent Equation	HIV Equation	Consent Equation	HIV Equation
Interviewer Effect				0.2125*** (0.020)		0.2120*** (0.020)
Years of Education	0.0020 (0.001)	0.0004 (0.000)	0.0023* (0.001)	0.0004 (0.000)	0.0023* (0.001)	0.0004 (0.000)
Wealth Category						
Poorest	-0.0187 (0.020)	0.0039 (0.005)	-0.0184 (0.019)	0.0039 (0.005)	-0.0184 (0.019)	0.0040 (0.005)
Poorer	-0.0563*** (0.020)	0.0078 (0.006)	-0.0553*** (0.020)	0.0078 (0.006)	-0.0554*** (0.020)	0.0079 (0.006)
Middle	-0.0747*** (0.021)	0.0017 (0.006)	-0.0769*** (0.021)	0.0017 (0.006)	-0.0769*** (0.021)	0.0018 (0.006)
Richer	-0.1139*** (0.024)	0.0021 (0.007)	-0.1169*** (0.023)	0.0021 (0.007)	-0.1171*** (0.023)	0.0023 (0.007)
Marital Status						
Currently Married	0.0241 (0.022)	0.0128 (0.008)	0.0235 (0.022)	0.0128 (0.008)	0.0237 (0.022)	0.0128 (0.008)
Formerly Married	-0.0382 (0.024)	0.0088 (0.008)	-0.0420* (0.024)	0.0087 (0.008)	-0.0418* (0.024)	0.0088 (0.008)
Had STD	0.0461 (0.029)	0.0093 (0.007)	0.0457 (0.028)	0.0094 (0.007)	0.0456 (0.028)	0.0094 (0.007)
Age at First Intercourse						
15 or Younger	0.0502** (0.022)	-0.0422 (0.035)	0.0485** (0.021)	-0.0254 (0.031)	0.0487** (0.021)	-0.0228 (0.030)
>15	0.0374* (0.022)	-0.0475 (0.034)	0.0356* (0.021)	-0.0334 (0.030)	0.0357* (0.021)	-0.0312 (0.030)
Had High Risk Sex	0.0531* (0.027)	-0.0275 (0.025)	0.0546** (0.027)	-0.0199 (0.022)	0.0544** (0.027)	-0.0184 (0.021)
Number of Partners						
1	0.0040 (0.022)	-0.0100 (0.007)	0.0033 (0.022)	-0.0100 (0.007)	0.0032 (0.022)	-0.0101 (0.007)
2+	-0.0064 (0.031)	-0.0178* (0.011)	-0.0068 (0.030)	-0.0180* (0.011)	-0.0069 (0.031)	-0.0180* (0.011)
Condom Last Intercourse	0.0001 (0.018)	0.0083* (0.004)	-0.0000 (0.018)	0.0082* (0.004)	-0.0000 (0.018)	0.0083* (0.004)
Would Care for HIV Relative	0.0221* (0.013)	-0.0007 (0.004)	0.0230* (0.012)	-0.0007 (0.004)	0.0229* (0.012)	-0.0007 (0.004)
Know Someone Who Died of AIDS	0.0163 (0.011)	-0.0009 (0.003)	0.0162 (0.011)	-0.0010 (0.003)	0.0161 (0.011)	-0.0009 (0.003)
Previously HIV Tested	0.0040 (0.018)	-0.0054 (0.007)	0.0046 (0.018)	-0.0054 (0.007)	0.0049 (0.018)	-0.0056 (0.007)
Smoker	-0.0103 (0.017)	0.0057 (0.004)	-0.0089 (0.017)	0.0057 (0.004)	-0.0089 (0.017)	0.0058 (0.004)

Language						
Akan	0.0474***	0.0042	0.0513***	0.0043	0.0511***	0.0044
	(0.017)	(0.004)	(0.016)	(0.004)	(0.016)	(0.004)
Ga	0.0766*	0.0009	0.0864**	0.0010	0.0858**	0.0010
	(0.045)	(0.013)	(0.044)	(0.013)	(0.044)	(0.013)
Ewe	0.0338	0.0122	0.0272	0.0122	0.0271	0.0123
	(0.045)	(0.009)	(0.045)	(0.009)	(0.045)	(0.009)
Nzema	0.0537	0.0253*	0.0484	0.0253*	0.0489	0.0254*
	(0.060)	(0.015)	(0.052)	(0.015)	(0.052)	(0.015)
Dagbani	-0.0201	-0.1142***	0.0187	-0.1387***	0.0186	-0.8298***
	(0.063)	(0.018)	(0.038)	(0.021)	(0.038)	(0.084)
Other	-0.0391	-0.0102	-0.0327	-0.0102	-0.0328	-0.0102
	(0.027)	(0.009)	(0.025)	(0.009)	(0.025)	(0.009)
Observations	4,955	4,955	4,955	4,955	4,955	4,955

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note to table 10: Controls for age, region, ethnicity, and religion are included in each regression equation but not shown. The model is a bivariate probit for HIV status and consent to a HIV test. Data are for men only. The first model uses interviewer fixed effects as the exclusion restriction. The coefficients from the interviewer effects are not shown in the table (see figure 9). The second model uses the two step random effects procedure where HIV consent is regressed on the X variables, along with the mean values for each interviewer (see table 9). The average error term for each interviewer is added to the predicted value of the interviewer means, and used as the exclusion restriction in the HIV regression. The final model is the Bayesian averaging procedure using the same exclusion restriction. Coefficients are obtained by restricting the value of RHO to its Bayesian estimate and running the bivariate probit. Source: DHS Ghana 2003 (men).

Table 11 Correlation Coefficient Estimates for Men in Ghana

Correlation Coefficient for Refused Consent	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
Fixed Effects Model	0.9452	-0.0514	0.9986		
Random Effects Model	0.9332	0.0197	0.9978	0.5603	0.9999
Bayesian Average Model	0.5927			0.3856	0.7184

Note to table 11: The table shows the estimated correlation coefficient between consent and HIV status for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Source: DHS Ghana 2003 (men).

Table 12 HIV Rate among Men who Refused to Test in Ghana

HIV Rate for Refused Consent	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
Fixed Effects Model	0.0000	0.0000	0.0000		
Random Effects Model	0.0000	0.0000	0.0000	0.0000	0.0016
Bayesian Average Model	0.0007	0.0005	0.0009	0.0003	0.0035
Imputation Model	0.0182	0.0160	0.2048		

Note to table 12: The table shows the estimated HIV prevalence rate among individuals who refused consent for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Source: DHS Ghana 2003 (men).

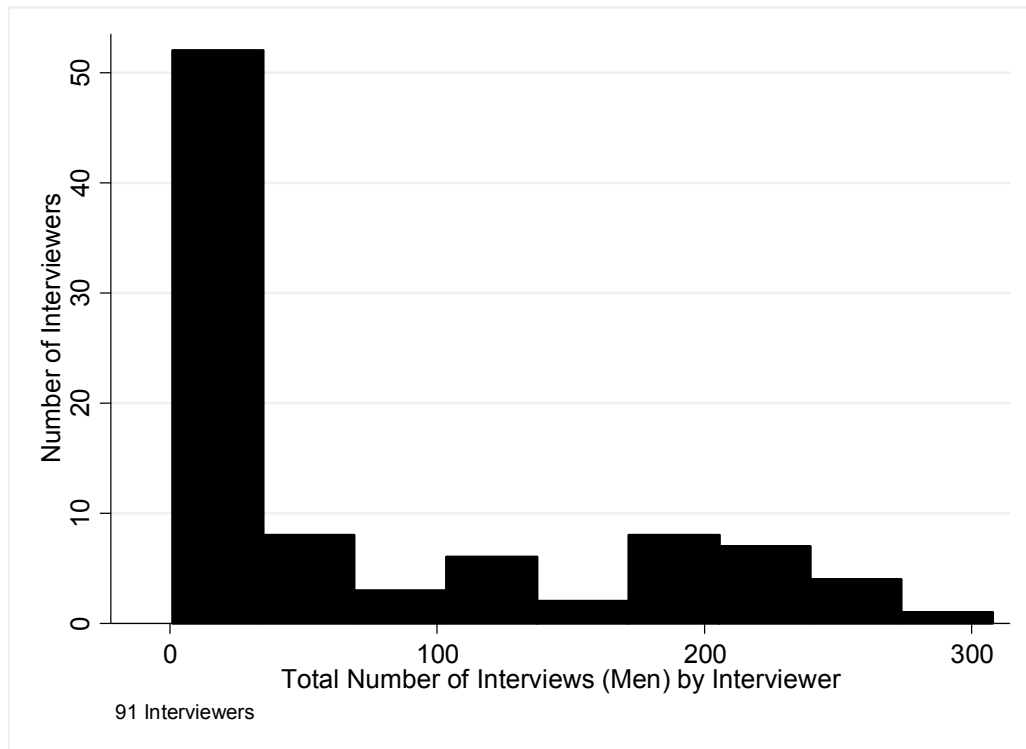
Table 13 HIV Rate among Men in Ghana

HIV Rate	Parameter Value	Analytic 95% CI		Bootstrap 95% CI	
All Men - Fixed Effects Model	0.0138	0.0105	0.0171		
All Men - Random Effects Model	0.0138	0.0105	0.0171	0.0116	0.0162
All Men - Bayesian Average Model	0.0139	0.0106	0.0172	0.0115	0.0163
Men with Non-Missing Data	0.0161	0.0120	0.0202		
Men with no contact - Imputation Model	0.0157	0.0137	0.0177		
All Men - Imputation Model	0.0164	0.0131	0.0197		

Note to table 13: The table shows the estimated HIV prevalence rate among the population for the fixed effects, random effects and Bayesian average models. Analytic standard errors are shown for the fixed effects and random effects models, with bootstrapped errors for random effects and Bayesian average models. Results from an imputation model are also shown, along with estimates only using those without missing data. Source: DHS Ghana 2003 (men).

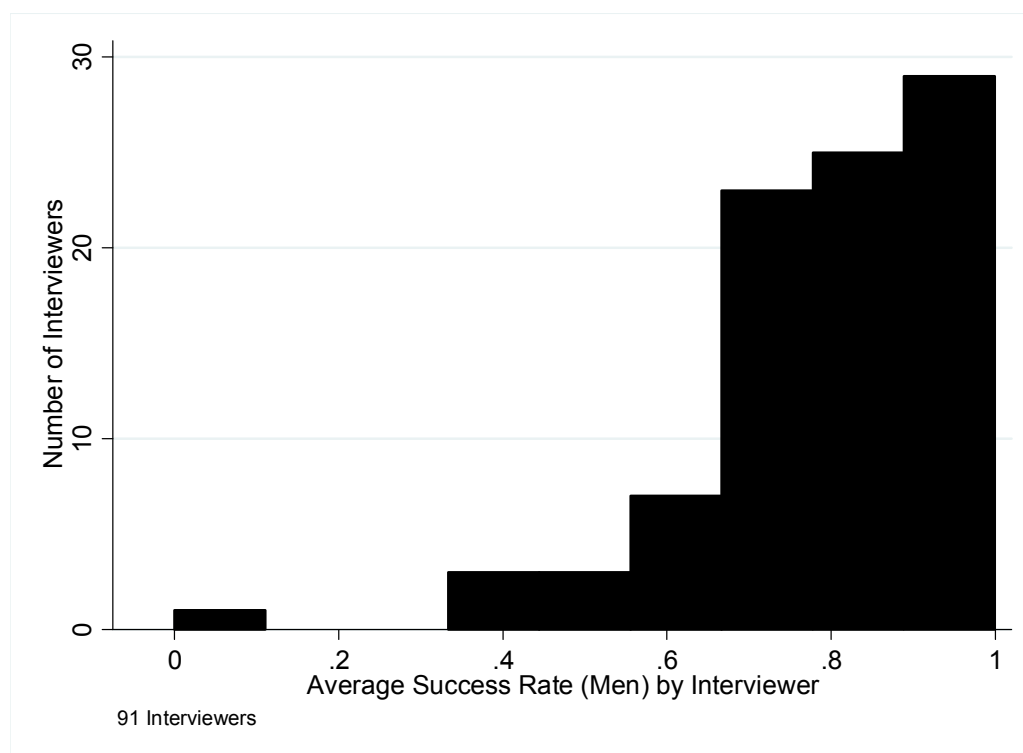
Figures

Figure 1 Number of Interviews by Interviewer Zambia 2007 (Men)



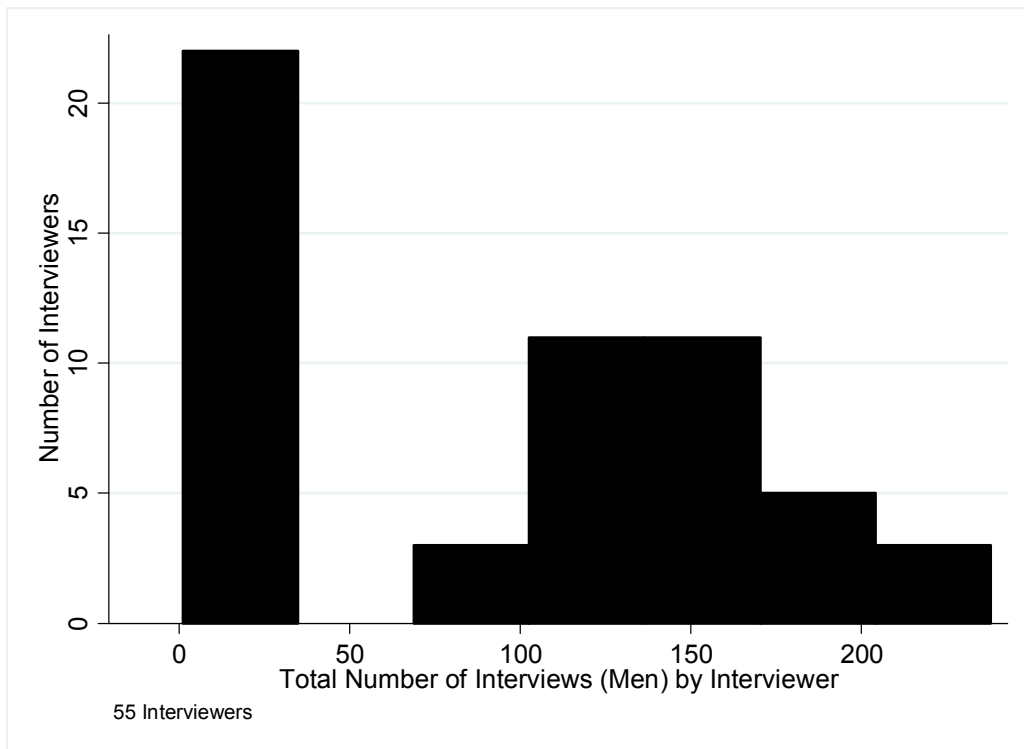
Note to Figure 1: Graph is at the interviewer level (one observation per interviewer).

Figure 2 Success Rate of Interviewers Zambia 2007 (Men)



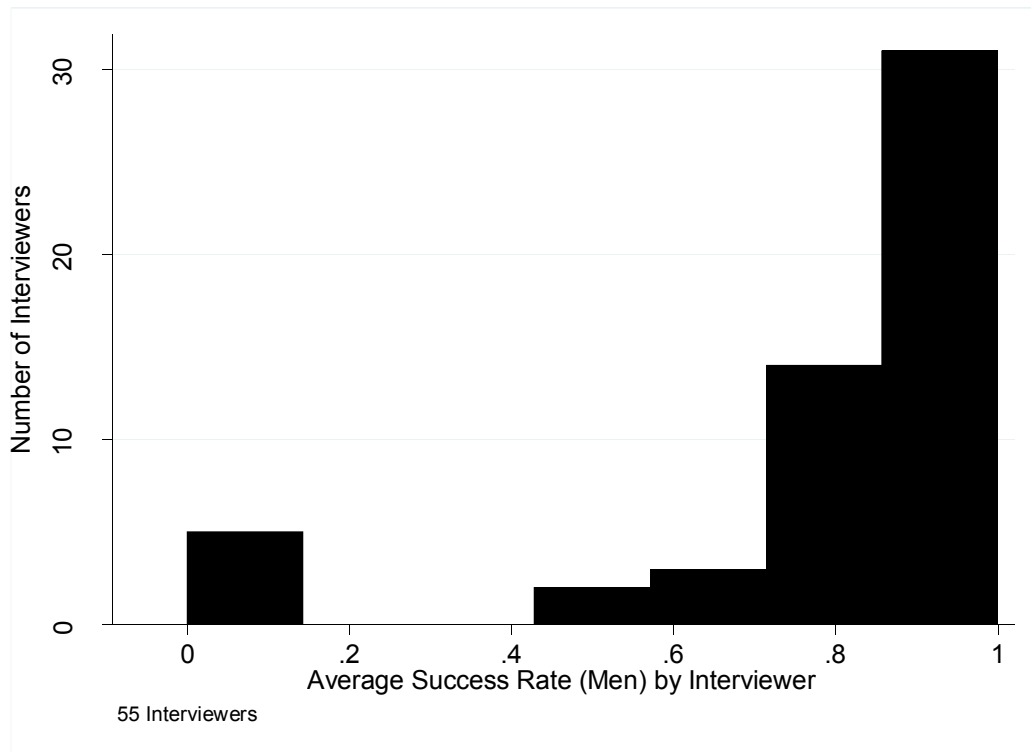
Note to Figure 2: Graph is at the interviewer level (one observation per interviewer).

Figure 3 Number of Interviews by Interviewer Ghana 2003 (Men)



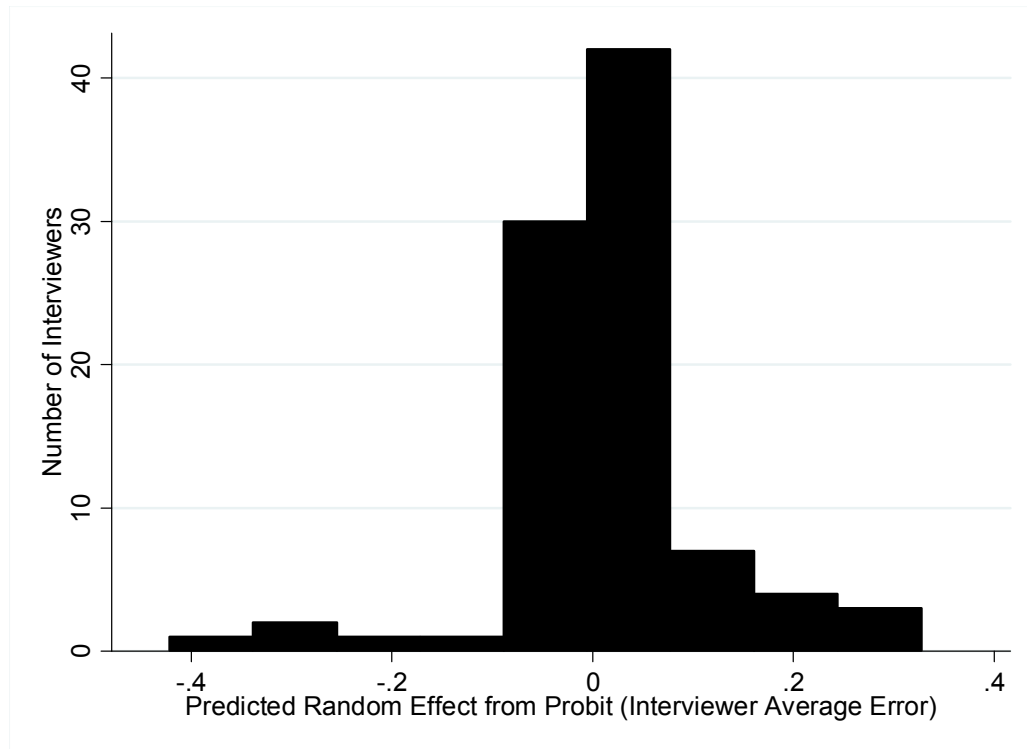
Note to Figure 3: Graph is at the interviewer level (one observation per interviewer).

Figure 4 Success Rate of Interviewers Ghana 2003 (Men)



Note to Figure 4: Graph is at the interviewer level (one observation per interviewer).

Figure 5 Random Effects from Probit for Consent Zambia 2007 (Men)



Note to Figure 5: Random effects are calculated as the average error term for each interviewer from a probit regression for HIV consent, including the interviewer average of each variable. See table 4. Graph is at the interviewer level (one observation per interviewer).

Figure 6 Interviewer Fixed Effect Coefficients Density Plot for Consent to HIV Test in Zambia (Table 5)

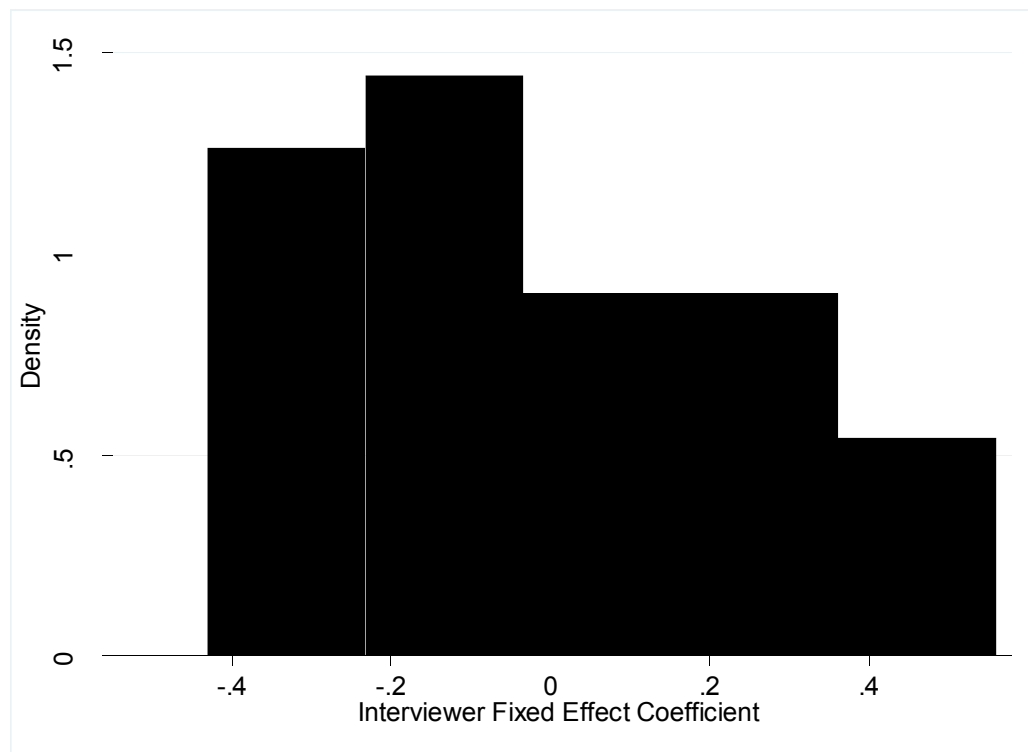
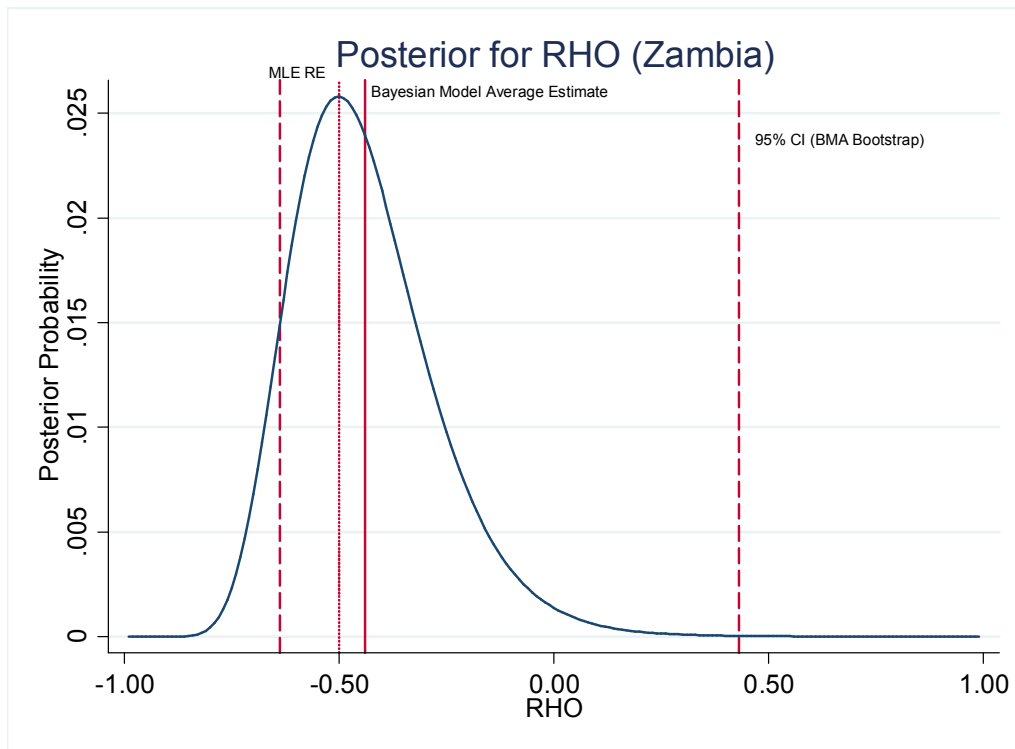
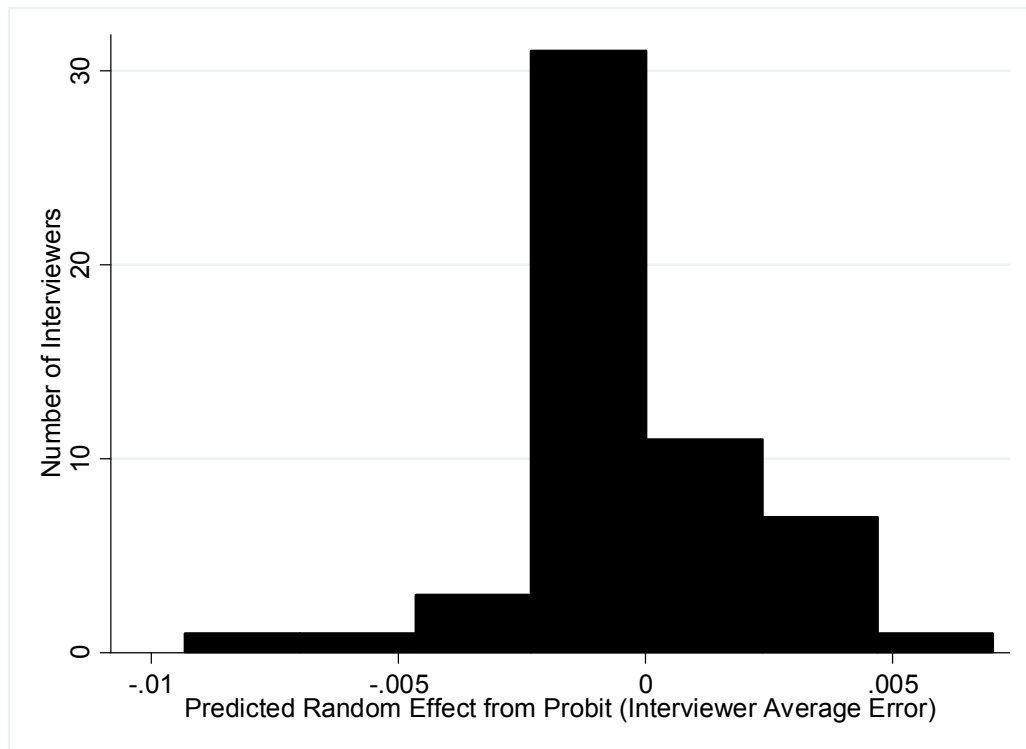


Figure 7 Bayesian Model Averaging Posterior for RHO Zambia 2007 (Men)



Note to Figure 7: Graph shows the posterior for RHO calculated using Bayesian averaging with interviewer random effects as the exclusion restriction. Bayesian averaging is performed over selection models with interviewer effects as the exclusion restriction. Random effects are obtained as the average error from a probit for consent. Also shown are the 95% bootstrapped confidence interval for the Bayesian average estimate, and the point estimate for the random effects maximum likelihood estimate for comparison. The bootstrap confidence interval is calculated using the empirical distribution of bootstrap estimates. Source: DHS Zambia 2007 (men).

Figure 8 Random Effects from Probit for Consent Ghana 2003 (Men)



Note to Figure 8: Random effects are calculated as the average error term for each interviewer from a probit regression for HIV consent, including the interviewer average of each variable. See table 9. Graph is at the interviewer level (one observation per interviewer).

Figure 9 Interviewer Fixed Effect Coefficients Density Plot for Consent to HIV Test in Ghana (Table 10)

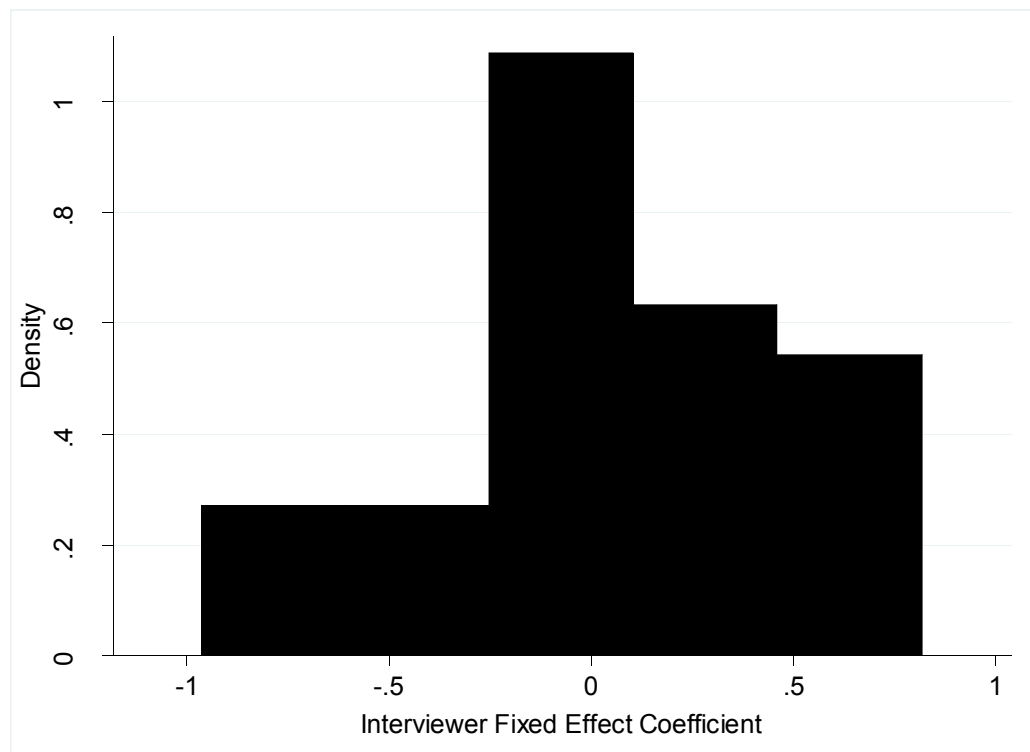
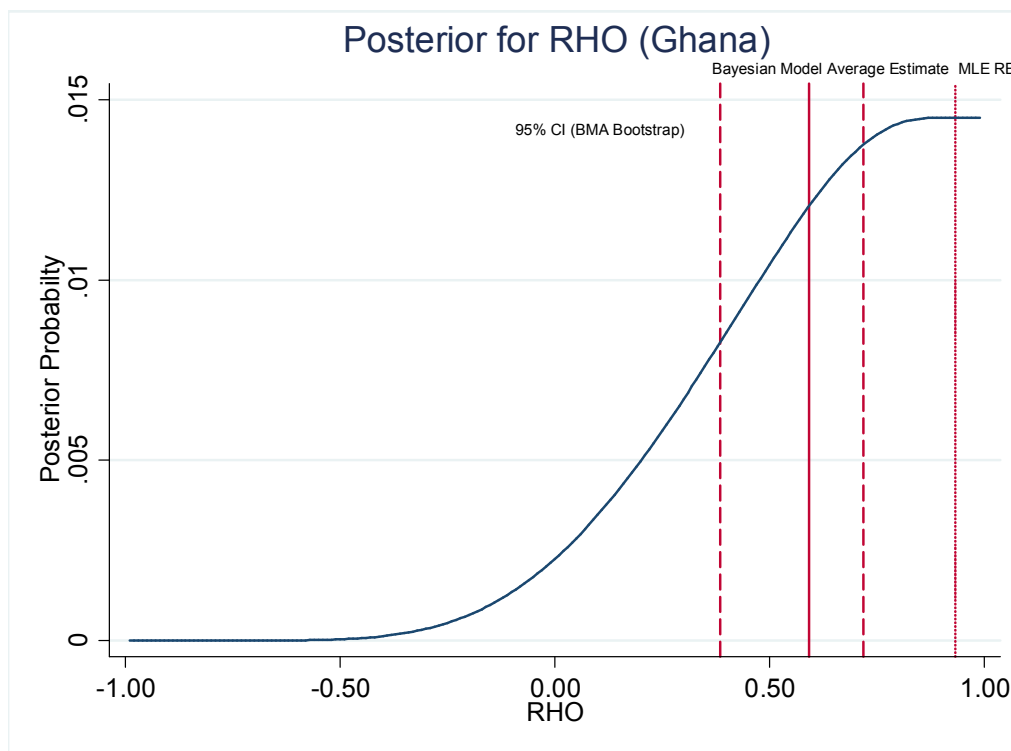


Figure 10 Bayesian Model Averaging Posterior for RHO Ghana 2003 (Men)



Note to Figure 10: Graph shows the posterior for RHO calculated using Bayesian averaging with interviewer random effects as the exclusion restriction. Bayesian averaging is performed over selection models with interviewer effects as the exclusion restriction. Random effects are obtained as the average error from a probit for consent. Also shown are the 95% bootstrapped confidence interval for the Bayesian average estimate, and the point estimate for the random effects maximum likelihood estimate for comparison. The bootstrap confidence interval is calculated using the empirical distribution of bootstrap estimates. Source: DHS Ghana 2003 (men).

Appendix

The concentrated likelihood function

The likelihood of the parameters (β_s, β_h, ρ) given the full data set (y, x, z) is

$$L(\beta_s, \beta_h, \rho) = P(y, x, z | \beta_s, \beta_h, \rho) \quad (9)$$

For a given ρ we can concentrate the likelihood function by setting the other parameters at their maximum likelihood values given ρ :

$$L_c(\rho) = P(y, x, z | \widehat{\beta}_s(\rho), \widehat{\beta}_h(\rho), \rho) \approx P(y, x, z | \rho) \quad (10)$$

In large samples the approximation to $P(y, x, z | \rho)$ will become exact as the maximum likelihood estimates of the other parameters are consistent. Using the concentrated maximum likelihood the problem is reduced to a one parameter model and we can carry out our Bayesian averaging approach with a prior over ρ alone.