4-2024

HIV estimation using population based surveys with non-response: A partial identification approach

Oyelola A ADEGBOYE

Tomoki FUJII
Singapore Management University, tfujii@smu.edu.sg

Denis H. Y. LEUNG
Singapore Management University, denisleung@smu.edu.sg

Siyu LI
Singapore Management University, siyu.li.2019@economics.smu.edu.sg

SMU ECONOMICS & STATISTICS

SMU SINGAPORE MANAGEMENT UNIVERSITY

HIV estimation using population based surveys with non-response: a partial identification approach

Oyelola A. Adegboye, Tomoki Fujii, Denis Heng-Yan Leung, Li Siyu

April 2024

Paper No. 04-2024

HIV estimation using population based surveys with non-response: a partial identification approach

Oyelola A. Adegboye[1*], Tomoki Fujii[2] , Denis Heng-Yan Leung[2], Li Siyu[2]

[1] Menzies School of Health Research, Charles Darwin University, Australia

[2]School of Economics, Singapore Management University, Singapore

[*] Corresponding author

April 11, 2024

Abstract

HIV estimation using data from the Demographic and Health Surveys (DHS) is limited by the presence of non-response and test refusals. Conventional adjustments such as imputation require the data to be missing at random. Methods that use instrumental variables allow the possibility that prevalence is different between the respondents and non-respondents, but their performance depends critically on the validity of the instrument. Using Manski's partial identification approach, we form instrumental variable bounds for HIV prevalence from a pool of candidate instruments. Our method does not require all candidate instruments to be valid. We use a simulation study to evaluate and compare our method against its competitors. We illustrate the proposed method using DHS data from Zambia, Malawi and Kenya. Our simulations show that imputation leads to seriously biased results even under mild violations of non-random missingness. Using worst case identification bounds that do not make assumptions about the non-response mechanism is robust but not informative. By taking the union of instrumental variable bounds balances informativeness of the bounds and robustness to inclusion of some invalid instruments. Non-response and refusals are ubiquitous in population based HIV data such as those collected under the DHS. Partial identification bounds provide a robust solution to HIV prevalence estimation without strong assumptions. Union bounds are significantly more informative than the worst case bounds without sacrificing credibility.

Key messages

- Partial identification bounds are useful for HIV estimation when data are subject to non-response bias
- Instrumental variables can narrow the width of the bounds, but the validity of an instrument variable is an untestable hypothesis
- This paper proposes pooling candidate instruments and creating union bounds from the pool
- Our approach significantly reduces the width of the worst case bounds without sacrificing robustness

1 Introduction

In sub-Saharan Africa, home to around 23 million people living with HIV [1], accurate measurement of the trends of important diseases such as HIV is essential for governments to design policies and aid programs. In the past two decades, national population-based surveys from the Demographic and Health Survey (DHS) system have become an important source for such measurement [2, 3]. A major challenge in using these data is the potential bias from missing data created by non-response. There is much evidence that the non-respondents may have patterns of outcome and/or behaviour that are very different from those of the rest of the population [4, 5].

One reason why non-response has garnered significant attention from researchers is the complexity of the problem [6]. Non-response is not a result of a single source or a well-defined situation, as it is widely recognized. Instead, the causes and processes that lead to non-response are diverse and often depend on multiple factors, including the surveyed population, the outcome's nature, and the survey's design and implementation. The most challenging aspect of this problem is that information about non-respondents is typically limited, making it challenging for surveyors to determine the reason behind a non-response [6] In the context of HIV survey, non-response arises primarily from two sources– non-contacts and refusals. The processes leading to these two types of non-responses are believed to be distinct. But for ease of discussion, we use these terms interchangeably. We return to distinguish them in the empirical study.

A primary concern when reporting HIV prevalence estimates using DHS data is potential bias resulting from non-response. Some relevant earlier works on non-response bias in HIV estimation using data from the DHS system include Garcia-Calleja et al. [3] and Larmarange et al. [7], who carried out multi-country surveys of response rates and evaluated non-response bias. Marston et al. [4] examined non-response bias in a nine-country study. They assumed non-response is non-informative and estimated the prevalence among the non-respondents by multiple imputation. Similarly, Mishra et al. [5] used a logistic regression to predict the HIV prevalence among the non-respondents under a non-informative non-response assumption in a twelve-country study. Reniers and Eaton [8] and Floyd et al. [9] corrected refusal bias in population surveys by using auxiliary longitudinal data. Their method relies on the assumption that the refusal behaviour in different populations are comparable. Reniers et al. [10], Bärnighausen et al. [11], Hogan et al. [12] adjusted non-response bias by a Heckman-type selection model [13], which allows non-response to be informative but requires the existence of a valid instrumental variable that satisfies the exclusion criteria of explaining non-response but not the outcome. Arpino et al. [14] constructed bounds based on the partial identification approach of Manski [15, 16]. Under this approach, the unknown quantity of interest can only be identified to within a set of bounds, whose width depends on the knowledge, or lack thereof, about the missing data. In this sense, the bounds are "worst case" bounds since no assumptions are made regarding the missingness process. Worst case bounds are often considered overly conservative in practice. Arpino et al. [14] used restrictions implied by the dynamics of HIV (ie., an infected person remains infected over time while an uninfected person cannot be infected earlier) and instrumental variables to narrow the width of the identification region.

Methods that use instrumental variables (IVs) allow the possibility that HIV prevalence is different between the respondents and those who refuse testing. However, valid instruments about the non-response mechanism are notoriously difficult to find. Furthermore, whether an instrument is valid is not a testable hypothesis. This paper aims to solve this conundrum. We espouse the view that, due to missing data, a study with missing data can never achieve as much as it would have had there been no missing data. This view departs from the conventional

wisdom that, with sufficient assumptions and modelling, that a study with missing data can be restored to the state as if there were no missing data save the fewer observations. Under the conventional perspective, unknown quantities of interest can be estimated using point estimates, or "point identified", with an adjustment to the reduced information, and then inferential tools such as confidence intervals and hypothesis tests can be carried out as usual. In our view, the uncertainty created by the missing data and our inability to pinpoint the exact causes of missingness must be embedded into the formulation of the analysis strategy.

Theoretically, we can take multiple candidate instruments if we do not know whether an instrument is valid. Indeed, in observational epidemiological studies that are subject to confounding or reverse causation bias, the use of genetic variants as proxies for environmentally modifiable exposures may lead to a hundred or more candidate instruments [17]. However, we do not know which instruments are valid among the instruments under consideration. We propose a two-stage modification of Manski's partial identification approach to solve this problem. Assume $s \geq a$, where a is the minimum number of valid instruments out of the L candidate instruments under consideration. For each candidate, we can use Manski's approach to form bounds. Then, even though we do not know the validity of individual instruments, the union of bounds using any set of $L-a+1$ individual candidates is guaranteed to correctly identify the quantity of interest. Following [18, 19, 20] that the intersection of bounds is non-empty for any set of valid instruments to eliminate the candidates whose bounds fail to overlap with the bounds of the majority of the candidates, we then take the intersection of the union bounds from all possible sets of $L-a+1$ instruments to form a new set of bounds. This step substantially narrows the bounds in some cases without sacrificing robustness. We carry out a simulation experiment to evaluate the proposed method. We then illustrate our method using data from the Zambia Demographic Health Surveys.

2 Method

We assume an outcome variable Y is measurable and bounded for each individual in the population of interest. Suppose we are interested in the population mean of Y, $\mathrm{E}(Y)$. In general, we may also be interested in $\mathrm{E}(Y|X)$ for some covariates X, but for brevity, we focus our discussion in the next two sections on estimating $\mathrm{E}(Y)$ since the treatment for the case with covariates is similar. Suppose a random sample of n is drawn from the population, and in this sample, Y is observed only in a subset of the sample. Let D be a binary variable such that $D = 1$ if Y is observed and 0 otherwise. Using the law of iterated expectations, we can write

$$\mathrm{E}(Y) = \mathrm{E}(Y|D=1)\mathrm{P}(D=1) + \mathrm{E}(Y|D=0)\mathrm{P}(D=0). \tag{1}$$

The sampling process identifies $\mathrm{E}(Y|D=1)$, $\mathrm{P}(D=1)$ and $\mathrm{P}(D=0) = 1 - \mathrm{P}(D=1)$ but there is no information on $\mathrm{E}(Y|D=0)$ unless we make strong assumptions about the joint distribution of Y and D. Let K_0, K_1 be, respectively, the lower and upper bounds of Y. Furthermore, write $\mu \equiv \mathrm{E}(Y)$, $\mu_{d\cdot} \equiv \mathrm{E}(Y|D=d)$. The worst case partial identification bounds [15] for μ are

$$(\mathrm{LB}, \mathrm{UB}) = (\mu_{1\cdot}\mathrm{P}(D=1) + K_0\mathrm{P}(D=0), \mu_{1\cdot}\mathrm{P}(D=1) + K_1\mathrm{P}(D=0)). \tag{2}$$

2.1 Bounds using instruments

The worst case bounds (2) are guaranteed to identify $\mathrm{E}(Y)$ by construction. However, they are often criticised for being too wide to be informative. The worst case bounds can be improved

if additional assumptions are made. Let V be an instrumental variable (IV) with discrete values $v \in \mathcal{V}$, such that,

$$\mathrm{P}(D = d|V = v_1) \neq \mathrm{P}(D = d|V = v_2), \tag{3}$$

and

$$\mathrm{P}(Y) = \mathrm{P}(Y|V = v_1) = \mathrm{P}(Y|V = v_2), \tag{4}$$

for $d = 0, 1$, all values $v_1, v_2 \in \mathcal{V}$ and $v_1 \neq v_2$. Write $\mu_{\cdot v} \equiv \mathrm{E}(Y|V = v)$ and $\mu_{dv} \equiv \mathrm{E}(Y|D = d, V = v)$. Since (4) implies $\mathrm{E}(Y|V = v) = \mathrm{E}(Y) = \mu$, it follows that [18], $\forall v \in \mathcal{V}$,

$$\mu_{1v}\mathrm{P}(D = 1|V = v) + K_0\mathrm{P}(D = 0|V = v) \leq \mu_{\cdot v} \leq \mu_{1v}\mathrm{P}(D = 1|V = v) + K_1\mathrm{P}(D = 0|V = v).$$

The inequalities imply

$$\begin{aligned}
&\mu \in \bigcap_{v\in\mathcal{V}} [\mu_{1v}\mathrm{P}(D = 1|V = v) + K_0\mathrm{P}(D = 0|V = v), \mu_{1v}\mathrm{P}(D = 1|V = v) + K_1\mathrm{P}(D = 0|V = v)] \\
\Rightarrow\quad &\mathrm{LB}_V \equiv \sup_{v\in\mathcal{V}} \{\mu_{1v}\mathrm{P}(D = 1|V = v) + K_0\mathrm{P}(D = 0|V = v)\} \leq \mu \qquad (5) \\
&\qquad \leq \inf_{v\in\mathcal{V}} \{\mu_{1v}\mathrm{P}(D = 1|V = v) + K_1\mathrm{P}(D = 0|V = v)\} \equiv \mathrm{UB}_V, \qquad (6)
\end{aligned}$$

where $(\mathrm{LB}_V, \mathrm{UB}_V)$ gives a set of IV lower and upper bounds for μ.

In practice, more than one instrument is usually used in a particular study [see, e.g., 17, 21]. Suppose there are L candidate instruments, and all we can assume is at least one of the L candidates is valid. Then, if some turn out to be invalid, (6) may fail to identify $\mathrm{E}(Y)$ for these instruments. To address this, suppose we create the following "union" bounds:

$$(\mathrm{LB}^{\mathrm{UN}}, \mathrm{UB}^{\mathrm{UN}}) = \bigcup_{\mathcal{V}_l, l=1,\cdots,L} (\mathrm{LB}_{V_l}, \mathrm{UB}_{V_l}) = (\inf_{v\in\mathcal{V}_l} \mathrm{LB}_{V_l}, \sup_{v\in\mathcal{V}_l} \mathrm{UB}_{V_l}). \tag{7}$$

It is trivial to see that $(\mathrm{LB}^{\mathrm{UN}}, \mathrm{UB}^{\mathrm{UN}})$ identifies $\mathrm{E}(Y)$ as long as at least one of the candidate instruments is valid. However, a simple examination of $(\mathrm{LB}^{\mathrm{UN}}, \mathrm{UB}^{\mathrm{UN}})$ reveals that as L increases, so will the width of $(\mathrm{LB}^{\mathrm{UN}}, \mathrm{UB}^{\mathrm{UN}})$. The wider a set of bounds, the less informative it is in identifying $\mathrm{E}(Y)$. Hence it would be of interest to eliminate among the L instruments, those that do not contribute to the identification of $\mathrm{E}(Y)$. To continue, we assume that the true number of valid instruments, s is known to satisfy $s \geq a \geq 1$ for some known a. Under this assumption, each subset of $(L - a + 1)$ instruments must contain at least one valid instrument. Hence, the union bounds formed by each subset is guaranteed to identify $\mathrm{E}(Y)$. For any two sets of bounds that both include $E(Y)$, their intersection must be non-empty, and also correctly identify $\mathrm{E}(Y)$. We therefore propose to find the intersection of all union bounds formed with any $(L - a + 1)$ instruments among the L instruments, because it will also identify $\mathrm{E}(Y)$ but be no longer than any of these union bounds.

Applying the bounds empirically incurs uncertainty, and this uncertainty can be incorporated in the form of confidence intervals. Suppose the theoretical lower and upper IV bounds for μ, denoted as $(\mathrm{LB}, \mathrm{UB})$, can be empirically estimated as $(\widehat{\mathrm{LB}}, \widehat{\mathrm{UB}})$. Constructing a confidence interval involves considering an approximate $b_0 \times 100$ percent interval for $(\widehat{\mathrm{LB}} - z_{(1-b_0)/2}\widehat{\mathrm{SE}}_{\mathrm{LB}}, \widehat{\mathrm{UB}} + z_{(1-b_0)/2}\widehat{\mathrm{SE}}_{\mathrm{UB}})$, where $z_{(1-b_0)/2}$ is the upper $(1 - b_0)/2 \times 100$ percentile of the standard normal distribution, SE represents standard error and $\widehat{\mathrm{SE}}$ its sample analogue. However, to address the width issue for μ [22], alternative bounds are suggested, $(\widehat{\mathrm{LB}} - C_n\widehat{\mathrm{SE}}_{\mathrm{LB}}, \widehat{\mathrm{UB}} + C_n\widehat{\mathrm{SE}}_{\mathrm{UB}})$, where C_n is determined through a specific equation involving the standard normal cumulative distribution function. The challenging analytical determination of standard errors leads to the utilization of bootstrapping, as proposed by Horowitz

and Manski [23] in this study. Bootstrap samples are obtained, and from these, the empirical standard errors $\widehat{\text{SE}}_{\text{LB}}$ and $\widehat{\text{SE}}_{\text{UB}}$ are derived to construct robust confidence intervals. Justifications and further details about the proposed bounds and confidence intervals are given in the Supplementary materials.

3 Simulation study

3.1 Simulation study 1

We use a simulation study to evaluate our proposed bounds (7). We assume the response Y is binary. We fix the values of s and L at 3 and 5, respectively. The instruments are all binary with a prevalence of 0.5 and mutually independent of each other.

We generate Y using a logistic model

$$\text{logitP}(Y = 1) = b_0 + b_{11}V_1 + \cdots + b_{1L}V_L, \tag{8}$$

where the coefficients $b_1 = (b_{11}, \cdots, b_{1L})^T$ give the association between the instruments and Y. A non-zero value of b_{1j} induces an association and therefore renders the instrument invalid. We use two different combinations for b_1: $b_1 = (\overbrace{0, \cdots, 0}^{s}, \overbrace{1, \cdots, 1}^{(L-s)})^T$; and $b_1 = (\overbrace{0, \cdots, 0}^{s}, \overbrace{4, \cdots, 4}^{(L-s)})^T$. For both situations, we assume without loss of generality the first s instruments are valid while the remaining $L-s$ are invalid. In the former, (4) is weakly violated by the invalid instruments while the violation of (4) is strong for the latter.

The non-response indicator D is generated using another logistic model

$$\text{logitP}(D = 1) = c_0 + c_{11}V_1 + \cdots + c_{1L}V_L + c_Y Y. \tag{9}$$

The coefficients $c_1 = (c_{11}, \cdots, c_{1L})^T$ give the association between each instrument and D. We consider two situations, (a) Strong instruments: $c_1 = (5, \cdots, 5)$ and (b) Strong + weak instruments: s coefficients are randomly given a value of 5 and the remaining $L - s$ are given a value of 0.5. The coefficient c_Y is used to model the association of D to the outcome Y, and hence selection bias. When $c_Y = 0$, then there is no selection bias when conditioned on the observed covariates. We consider two choices of $c_Y = -0.1||c_1||$ and $-0.3||c_1||$, where the symbol $||\cdot||$ stands for the sum of the coefficients $c_{11}, ..., c_{1L}$. We use negative association to reflect that in practice, we expect those who are HIV positive are less likely to have an HIV test. These two values for c_Y correspond to mild to moderate selection bias. We use c_0 to calibrate the average non-response rate, $1 - \text{E}(D = 1)$, to be 0.1 and 0.3 over the simulations.

Since Y is binary, the bounds for Y are $(K_0, K_1) = (0, 1)$. Throughout the simulation study, a sample size of $n = 1000$ observations is used. We use 1000 simulation runs for each combination of parameters. Confidence intervals are approximated using the method described in the Supplementary materials. These confidence intervals require estimates of the standard errors of the bounds, which can be carried out using bootstrapping.

A standard approach to adjust HIV prevalence estimates for survey non-response is by imputation [24]. Using imputation, the missing outcomes are imputed using predicted prevalence based on observed information such as demographic, socio-economic and behavioural variables from those who were tested. We compare this method to the partial identification method. For the imputation method, we use all the observed variables in the simulation study, i.e., the instruments. For partial identification, we used the worst case bounds that do not make any assumptions, and also the method proposed in this article.

Table 1 gives the simulation results. Each combination of parameters corresponds to four rows of results. The first row shows the proportion of times, out of 1000 simulations, the approximate 95 percent confidence intervals include E(Y). The second row gives the lower confidence limits, averaged over 1000 simulations. The third row gives the upper confidence limits, averaged over 1000 simulations. The fourth row gives the average width of the confidence intervals.

When non-response probability is 0.1 and selection bias is mild, 95% confidence intervals using all three methods have high probabilities of capturing E(Y). Using imputation naturally leads to much narrower confidence intervals. Between the partial identification bounds, the IV bounds proposed in this paper produce a much narrower confidence interval but at the expense of not capturing E(Y) in finite samples.

In all other situations, using imputation leads to grossly biased confidence intervals that fail to capture E(Y) in almost all simulation runs. Recall that E(Y) is calibrated to be at 0.15 in all simulations so the imputation confidence intervals underestimate the true prevalence. The advantage of the IV bounds confidence intervals over the worst case confidence intervals mirrors those when non-response probability is 0.1 and selection bias is mild. Additional simulations have been carried out. The results are given in the Supplementary materials. The conclusion from the additional simulations is similar to those presented here.

In practice, selecting the value of s may be challenging. To address this, we also carried out a sensitivity analysis that presents bounds for different values of s as a sensitivity parameter. The results show the following general conclusions. First, as long as the assumed number of valid instruments a is no less than $s \geq 1$, the proposed method produces bounds and associated CIs that capture the unknown HIV rate. Second, when the assumption $s \geq a \geq 1$ is violated but the invalid instruments are only weakly invalid, the performance of the proposed method is still satisfactory. When $s \geq a \geq 1$ is violated and the invalid instruments are strongly invalid, the proposed method does not perform well. Third, a smaller value of a is more robust to the violation of the assumption $s \geq a \geq 1$. Fourth, a larger value of a gives narrower bounds but as pointed out above the reduction comes at a price when the assumed number of valid instruments a exceeds the actual number s, in which case the shorter bounds fail to capture the unknown HIV rate. Hence, a balance needs to be struck between the two goals. Obviously, in practice, we must have certain confidence in the validity of the candidate instruments before they should be included into consideration for creating the bounds. To conserve space, the full results are given in the Supplementary Materials.

3.2 Simulation study 2

In this section, we report details of a second set of simulations that allows comparison between the proposed method and methods from Marra et al. [25] and Jiang and Ding [26]. The simulations follow the approach of Clark and Houle [27] by using data with a structure similar to a real DHS survey. We use the 2007 Zambia DHS men sub-sample as the basis of our simulation setup.

The relevant individuals in the survey are men eligible for individual surveys. In the 2007 Zambia DHS, eligible individuals were first approached for the individual surveys. Those who were contactable and present at the individual surveys were then asked to participate in HIV testing. The eligible individuals can be classified into one of three groups: (a) those who were absent for the individual surveys and not tested, (b) those who participated in the individual surveys but refused to be tested, and (c) those who participated in the individual surveys, agreed to be tested and with a valid test results. For those in groups (a) and (b), their HIV test results are absent. For our simulation setup, HIV results are generated through a three-

stage process: (1) contact for individual surveys, (2) consent to HIV test, and (3) test results among the tested.

Details of the simulations are given in Section S.4 of the Supplementary Materials. Here, we briefly describe the simulation setup. We fix a sample size of $n = 7000$. Then, we simulate observations with a composition of age, rural residence and geographical region similar to that of the original survey. Following Marra et al. [25], we randomly generate an interviewer IV with 30 interviewers. We additionally generate 3 binary IVs: V_2 V_3, V_4 each with a prevalence of 0.5 in the samples.

To simulate data, we mimic the three-stage process described earlier by using three equations: a contact equation, a selection (consent) equation and an outcome equation. GPS coordinates of each of the 319 clusters are obtained and used to simulate spatial correlations in HIV rates. Parameter values used in these equations are obtained by fitting similar equations to the actual Zambia 2007 DHS data and calibrated to create simulated contact rate, refusal rate and HIV prevalence similar to those in the Zambia 2007 DHS (10%, 25% and 20%, respectively).

In the simulations, we consider two different situations of IV strength (Weak) and (Strong). Among the four IVs, we fix V_4 as a valid IV. We consider four different cases in terms of the number of invalid instruments: (1) three (V_2, V_3, and Interviewer IV) are invalid are invalid, (2) two (V_2 and V_3) are invalid, (3) one (V_2) is invalid, and (4) none is invalid. We generate random N(0,1) values to represent interviewers' differential persuasiveness in eliciting acceptance to HIV test.

For the method of Marra et al. [25], we use only interviewer as the IV. We try a selection of representative copulas: Normal, Frank, Clayton rotated 90 degrees, and Clayton rotated 270 degrees and then choose the best among them based on AIC. The effect of region is modelled using a Markov random field smoother (see Marra et al. [25]).

The method proposed here always uses all $L = 4$ IVs and assumes at least $s = 2$ of them are valid. Under case (1), the assumption is violated. Therefore, the four cases allow us to test the sensitivity of the method to the choice s. For the Jiang and Ding [26] method, all observations with $C = 0$ (non-contact) are assumed to have the same prevalence as those tested.

We consider 8 scenarios defined by the combination of the interviewer IV strength (weak/strong) and the number of invalid IVs (cases (1)–(4)). For each scenario, we use 1000 simulations to compare the following methods: (1) standard imputation; (2) the selection method of Marra et al. [25] assuming a Normal copula; (3) the selection using the best of the four copulas; (4) the worst case bounds; (5) the bounds from Jiang and Ding [26] and (6) the bounds proposed in this paper. Table 2 shows the summary statistics of each set of simulations. The first column shows the average true HIV prevalence; the second column shows the "observed" after removing the non-contacts and refusals; the third column shows the consent rates among those contacted; the fourth column shows the F-statistic for interviewer effect; and the last column shows the contact rates. Table 3 gives the average 95% lower and upper confidence intervals using each of the 6 methods. Table 4 shows the widths of the 95% confidence intervals using the 6 methods.

Comparing the results from Table 3 to the true prevalence rates in Table 2, it is obvious that imputation grossly underestimates the true rates. The Marra selection method, whether using Normal copula (the correct one under the setup of the simulation study) or the best, adjusts the observed prevalence upwards. When the interviewer IV is invalid (case (1)), the Marra confidence intervals give over-estimates of the true prevalence as expected. For the remaining scenarios, the average Marra lower confidence limits are near the true rates. So, the method still tends to adjust more than required. This is not surprising since in

the setup of the simulations, selection is based on the effectiveness of each interviewer (see the selection equation S.12 in Supplementary Materials where interviewer is represented by effectiveness). However, in practice, since this effectiveness is unobserved, the interviewer ID is just a surrogate of the interviewer's (unobserved) effectiveness.

All the partial identification methods produce confidence intervals that adjust upwards the observed prevalence. The bounds methods are designed to capture true prevalence. However, their lower confidence limits are all quite small, which is an artefact of assuming all non-tested as HIV negative in creating the lower bounds. In the simulations, there is little difference between JD and the worst case bounds. As expected the widths of the bounds confidence intervals are in the following orders: worst case $\geq$ JD $\geq$ union (Table 4). Comparatively, the Marra confidence intervals are much shorter.

This simulation study highlights that there is no single perfect method for bias correction when there are non-responses: the Marra method using interviewer IV is subject to the inability to observe the innate effectiveness of each interviewer; the partial identification methods are less sensitive to this issue, but tend to produce results that are too conservative (lower confidence limits that are even much lower than the observed prevalence).

4 Empirical application

The primary data source for this study is from three DHS. We first use the 2007 Zambia DHS to illustrate the method. We then apply the methods to two other DHSs (Malawi and Kenya).

4.1 HIV prevalence in Zambia

The 2007 Zambia DHS is the fourth survey in the Zambia DHS series and provides population-level health estimates, including data useful in monitoring and evaluating population, health, and nutrition programs.

A total of 7,969 households were selected for the 2007 Zambia DHS, of which 7,326 were occupied. The shortfall was largely due to households that were away for an extended period of time and structures that were found to be vacant at the time of the interview. Of the occupied households, a total of 7146 were successfully interviewed. The interviews collected basic demographic information (e.g., age, sex), socio-economic status (e.g., educational attainment) as well as basic household characteristics (e.g., household possessions and dwelling characteristics).

In the interviewed households, 7406 females were eligible for interview and HIV testing, while the number of eligible males was 7146. The individual interviews collected information such as work and background characteristics, marriage and sexual activities, and awareness and attitudes towards HIV. In the women's interviews, additional questions about reproductive history, child health, and nutrition were asked.

Of the women and men eligible for individual interviews, 1695 (22.9%) of the women and 1983 (27.7%) of the men refused or did not complete an HIV test. The primary reason for non-response among eligible men was the failure to find individuals at home despite repeated visits to the household, followed by refusal to be interviewed. The substantially lower response rate for men reflects the more frequent and longer absence of men from the households.

The interviews in the 2007 Zambia DHS were carried out by 12 teams made up of 12 supervisors, 12 editors, 36 female interviewers, and 36 male interviewers. Each team consisted of one supervisor, one female field editor, one laboratory technician, three female interviewers, and three male interviewers. The interviews and questionnaires were translated from English

into one of seven major local language groups: Nyanja, Bemba, Kaonde, Lunda, Lozi, Tonga, and Luvale.

The observed prevalence of HIV positive among the cases, with results stratified by age, are given in Table 5. Even though in this study, the proportions of non-response is modest, we shall see that using instruments still improves inferences in some cases.

The final list of instrumental variables we use are: iv.lan (whether the language used in the questionnaire or interview is the same as the respondent's language, yes *vs.* no), iv.firstday (whether the interview was conducted on the first day of the interviews, yes *vs.* no), iv.interviewer (number of interviews the interviewer has performed, < 50, $50 - 100$, $100 - 200$, > 200), iv.mon (whether the interview was carried out during a month of harvest or planting, yes *vs.* no), iv.doa (whether the respondent knows someone who has died of AIDS, yes *vs.* no).

It is well known that the validity of an instrument (4) is an untestable hypothesis. Nevertheless, we can determine whether an instrument is strong by evaluating (3). Table 6 shows chi-square tests between non-response and the candidate instrumental variables we consider; all tests are highly significant.

We assume $a = 3$, that is, at least three out of the five candidates are valid. In any survey, such as the 2007 Zambia DHS, non-response and the potential for an associated bias are always concerns. The standard procedure is an imputation analysis on those who are not tested to adjust for potential biases [28]. The individuals in the survey can be classified into one of three groups: (a) those who participated in the household and individual surveys and were tested (b) those who participated in the household and individual surveys but were not tested, and (c) those who only participated in the household surveys. For those in groups (b) and (c), their HIV test results are absent.

For individuals in groups (b) and (c), their probability of HIV is predicted based on multivariate models using data from those who were tested. A logistic regression model is used to calculate HIV probability separately for groups (b) and (c). For group (b), the variables used in the model include the following household survey variables: age, education, wealth quintile, residence, and geographic region, as well as the following variables from the individual survey: marital union, current work status, media exposure, religion, sexually transmitted infections (STIs) or STI symptoms in past 12 months, cigarette smoking/tobacco use, age at first sex, number of sex partners in past 12 months, higher-risk sex in past 12 months, condom use at last sex in past 12 months, and willingness to care for a family member with AIDS. Prediction for group (c) uses only the household variables. The models are used to impute HIV statuses for individuals in groups (b) and (c) and the results are combined with those in group (a) to form adjusted HIV prevalence estimates for the population.

For all estimates, the data are weighted by survey weights. For individuals in group (a), HIV weights were used, for individuals in group (b), the individual survey weights were used and for those in group (c), household survey weights were used.

We compare adjustments using standard imputation with those using partial identification bounds. For partial identification bounds, we report results based on the worst case bounds as well as the instrumental variable bounds. For brevity, we only report the 95% confidence intervals (CIs) in Table 7a-b. We also include confidence intervals based on the observed, unadjusted prevalence among the tested individuals. The results are stratified by gender and by age groups.

We hereafter focus the discussion on men's results, the women's results exhibit similar patterns. The imputation method uses models based on data from the tested, and hence implicitly, it assumes that conditioned on the covariates used in the models, a non-tested individual has the same propensity of HIV as a tested individual. This fact is borne out in the

95% CIs using imputation. All of them include the corresponding observed prevalence among the tested in Table 5, and they are all very similar to the corresponding CIs based on the unadjusted prevalence estimates. Due to the additional observations used, each imputation CI is never wider than its unadjusted counterpart. Both have relatively short widths due to the large sample sizes in this study.

4.2 HIV prevalence in Malawi and Kenya

The approaches described above were applied also to the 2004 Malawi DHS (MDHS) and the 2003 Kenya DHS (KDHS). These surveys have a design comparable to MDHS and KDHS have 2007 Zambia DHS for our purpose. Malawi and Kenya have lower HIV prevalence (based on unadjusted results). We use the same set of instruments as Zambia, except that the definitions of iv.interviewer and iv.mon are slightly different because of the differences in the number of interviews conducted by interviewers and differences in the agricultural seasons.

The results for Malawi and Kenya are broadly comparable to those for Zambia presented in Section 4.1. The imputation method gives very similar results to the unadjusted results. The partial identification bounds always give confidence intervals that are much wider. Between the two partial identification methods, the worse case is always less precise than the method proposed in this paper. The improved precision of the proposed method comes from a big reduction of the upper confidence interval, ranging from about 10% to 30%. Further details about the data, implementation and results are provided in Supplementary materials S.6 and S.7,

5 Discussion

Existing studies on refusal bias in the estimation of HIV prevalence typically either provide some evidence of the existence of the bias or try to correct for the bias by making some (often strong) behavioural assumptions about the subjects. In this paper, we have instead derived plausible lower and upper bounds for HIV prevalence under mild and intuitive assumptions. This approach is potentially useful because it is often difficult to validate or falsify an underlying assumption. Furthermore, it shows that a carefully designed and implemented localised study may also be helpful for understanding the magnitude of non-response bias.

Partial identification approach using instruments has been widely used in the fields of Social Sciences and Economics, though rare in Epidemiology and Public Health. As with other methods that exploit instruments, the key to the success of this approach is the validity of the instruments used to create the bounds. However, it is well known that the exclusion restriction assumption is a non-testable hypothesis. This paper offers a novel and simple solution to this challenge by taking multiple candidate instruments. If at least one instrument in the pool of candidates is valid, the proposed approach creates bounds that, in large samples, partially identify the true prevalence. The approach offered in this paper is especially useful for practitioners because normally, there are multiple variables, e.g., interviewing process, interviewer characteristics, etc., that are candidates to be considered as instruments, and yet there is no way to determine which one(s) is(are) valid. Using a large pool increases the chance of finding at least one that is valid but at the same time, induces the possibility of including invalid ones. The proposed method solves this conundrum.

We examine HIV prevalence between genders and across different age groups. In evaluating the proposed imputation method for instrumental variables, the study employed factors like language, interview timing, interviewer experience, month of interview, and awareness of AIDS-related deaths. The results demonstrated that the imputation method yielded similar

outcomes to unadjusted results across various scenarios. Furthermore, we compare various bounds based on a variety of methods, including Marra et al. [25], Manski's[22] worst case bounds, and Jiang and Ding [26]. Previous studies have suggested that variables related to the data collection process may be used as instruments because they affect the response probability but are unlikely to have a direct effect on the outcome [12, 29]. For example, an experienced interviewer or an interviewer of a similar age as the interviewee may have a better chance of eliciting a positive response. Furthermore, whether the language of the interview or questionnaire is the same language as the interviewee may affect the response rate. It has also been argued that the timing of the first interview attempt coincides with the economic cycle and affects the probability of finding the interviewees at home. Individuals selected to be interviewed on the first day of the interviews within a cluster of households will also have more chances to be contacted even if they are not at home, giving rise to a higher response probability. Finally, we also consider a variable based on the individual's attitude to HIV. The current literature finds that more negative attitudes are associated with refusal of an HIV test or never having had an HIV test in sub-Saharan Africa [30, 31].

Our proposed method is similar to that proposed in Kang et al. [32] for estimating causal effects when some instruments are possibly invalid. Kang et al. [32] also considers a union method, but their context and process differ from the present paper. In their paper, the goal is to obtain a confidence interval of some causal effect. They also assume a pool of N instruments with no more than s^* valid (in our notations). For each set of $s^* - 1$ instruments, they form a confidence interval of the causal effect. They then take the union of confidence intervals over all $\binom{N}{s^*-1}$ sets of instruments. On the contrary, our method first creates the partial identification bound using each instrument, then finds the union of bounds from every set of $N - s^* + 1$ instruments. In Kang et al. [32], the interval is narrowed by pretesting and eliminating possibly invalid instruments. In our paper, no tests are used. Instead, we take the intersection of the $\binom{N}{N-s^*+1}$ union bounds.

Using partial identification, the corresponding CIs are much wider than those using imputation. The much wider CIs using partial identification reflect the uncertainties we have about the actual HIV status of the non-tested individuals. The lower limits of the partial identification CIs are also, in general, quite a bit smaller than the corresponding imputation lower limit. The reason is that a lower partial identification bound is derived by assuming all non-tested individuals are HIV negative (for a given value of instrument for the IV bound), whereas imputation assumes the non-tested are the same as the tested, given the covariates. Similarly, the upper limits of the partial identification CIs are much higher than those given by imputation since the upper partial identification bounds result from assuming all non-tested are HIV positive (for a given value of instrument for the IV bound).

The most significant difference between the partial identification CIs and the imputation CIs lies in those situations where the observed prevalence among the tested is low. For example, in the male aged 15-19 group. The upper limit of the imputation CI is 0.045, against the upper limits of 0.32 and 0.278, respectively, for the worst case and instrumental variable partial identification CIs. The reason for the very low upper limit for the imputation CI is that it assumes those non-tested also have similarly low prevalence as the tested individuals. In contrast, the partial identification approach allows for the possibility that even if a moderate proportion of the non-tested are actually HIV positive, the prevalence would change significantly upwards.

Between the two partial identification methods, the worst case scenario makes no assumptions and the resulting CIs are wider than those derived using the instrumental approach proposed in this paper. Since the width of a CI gives its precision, the method proposed here is always more precise than the worst case CIs. In some cases, such as males aged 35-39, the

gain in precision approaches 30%.

Three other observations are worth noting. First, as expected, all CIs using the proposed method have a larger lower confidence limit than the corresponding worst case CIs. A second observation is the narrowing in widths in the CIs in the proposed method mainly comes from a much smaller upper limit than the corresponding worst case CIs. This advantage is brought about when the population is stratified by different levels of a valid instrument. if the proportion of tested individuals is higher at a particular level, the more precise information from such a group can be used to infer the HIV prevalence of the entire population. Finally, it is also notable that the gains in the length of a confidence interval from the IV bound—or the difference in the length of confidence intervals between the worst case and IV bounds—are much higher for men than for women in all three countries. This gender difference may be due to males being more sensitive to the interviewer's characteristics than females in agreeing to conduct HIV tests. While this result is not necessarily applicable to other contexts, it does show the relevance of the respondent and interviewer characteristics and potential gains from adopting our approach.

There are two cautionary notes to be made about the proposed method. First, it is advisable not to include too many instruments, particularly highly dubious ones. The goal of the proposed method is to create bounds that identify the prevalence robust to invalid instruments. However, if we add mostly invalid instruments, L will increase without a corresponding increase in a. This will result in an increase in the number of instruments, $L - a + 1$, leading to a wide confidence bound in the proposed method. Therefore, a balance must be struck between how many and what instruments should be included as candidates. Obviously, we should include as many as needed so we have comfort that some among the pool of candidates would be valid. Our simulations and empirical examples suggest that just a few candidate instruments would suffice. Second, the proposed method is not immune to the problem of weak instrumental variables. A weak instrumental variable is one that is not informative about the non-response process. Instrumental variable bounds based on weak instruments may be very wide and not much different from the worst case bounds. Therefore, we must also be judicious about the choice of instruments.[1] Fortunately, the strength of an instrument is a testable hypothesis. We demonstrated how this can be done in the empirical study (Table 6).

In conclusion, the proposed approach is useful for providing HIV prevalence estimates in population-based surveys, where non-response is a ubiquitous phenomenon and little is known about its causes.

acknowledgements

We acknowledge ORC Macro for granting us access to the DHS data sets.

conflict of interest

None declared

[1]A related point is how we define IVs. In many empirical settings, one could choose to have two binary instruments v_1 and v_2 or the combination of the two $v = 2v_1 + v_2$, which contain the same information as v_1 and v_2 combined. We would recommend to use v only when the validity of v_1 and v_2 is (likely to be) the same. When they are valid, v would be more useful than having v_1 and v_2 separately, as the former would give more variations in $P(D|V)$. If both of them are invalid, we reduce the proportion of invalid instruments. However, treating v_1 and v_2 as separate instruments would be more desirable if their validity is different.

data availability

This study is based on publicly available data, the Zambia DHS, Malawi DHS and Kenya DHS obtained from `www.dhsprogram.com`. The program used in this simulation study can be obtained from `https://github.com/oyeadegboye/partialindentification`

References

[1] WHO. Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report 2011. WHO, Geneva; 2011.

[2] Boerma JT, Ghys PD, Walker N. Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. Lancet 2003;362:1929–1931.

[3] Garcia-Calleja JM, Gouws E, Ghys PD. National population-based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. Sexually Transmitted Infections 2006;82(Suppl 3):iii64–iii70.

[4] Marston M, Harriss K, Slaymaker E. Non-response bias in estimates of HIV prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys. Sexually Transmitted Infections 2008;84 (Suppl 1):i71–i77.

[5] Mishra V, Barrere B, Hong R, Khan S. Evaluation of bias in HIV seroprevalence estimates from national household surveys. Sexually Transmitted Infections 2008;84 (Suppl 1):i63–i70.

[6] Adegboye OA, Fujii T, Leung DH. Refusal bias in HIV data from the Demographic and Health Surveys: Evaluation, critique and recommendations. Statistical Methods in Medical Research 2020;29(3):811–826.

[7] Larmarange J, Vallo R, Yaro S, Msellati P, Méda N, Ferry B. Estimating effect of non response on HIV prevalence estimates from Demographic and Health Surveys. Working Paper du CEPED, number 03, UMR 196 CEPED. Université Paris Descartes, INED, IRD, Paris; 2009.

[8] Reniers G, Eaton J. Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. AIDS 2009;23:1–9.

[9] Floyd S, Molesworth A, Dube A, Crampin AC, Houben R, Chihana M, et al. Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias. AIDS 2013;27:233–242.

[10] Reniers G, Araya T, Berhane Y, Davey G, Sanders EJ. Implications of the HIV testing protocol for refusal bias in seroprevalence surveys. BMC Public Health 2009;9(163).

[11] Bärnighausen T, Bor J, Wandira-Kazibwe S, Canning D. Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. Epidemiology 2011;22(1):27–35.

[12] Hogan DR, Salomon JA, Canning D, Hammitt JK, Zaslavsky AM, Bärnighausen T. National HIV prevalence estimates for sub-Saharan Africa: controlling selection bias with Heckman-type selection models. Sexually Transmitted Infections 2012;88 (Suppl 2):i17–i23.

[13] Heckman JJ. Sample selection bias as a specification error. Econometrica 1979;47:153–161.

[14] Arpino B, De Cao E, Peracchi F. Using panel data for partial identification of human immunodeficiency virus prevalence when infection status is missing not at random. Journal of Royal Statistical Sciety A 2014;177:587–606.

[15] Manski CF. Anatomy of the selection problem. Journal of Human Resources 1989;24(3):343–360.

[16] Manski CF. Nonparametric bounds on treatment effects. American Economic Review 1990;80(2):319–323.

[17] Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 2008;27(8):1133–1163.

[18] Manski CF. Partial Identification of Probability Distributions. New York: Springer-Verlag; 2003.

[19] Chernozhukov V, Lee S, Rosen AM. Intersection bounds: estimation and inference. Econometrica 2013;81(2):667–737.

[20] Windmeijer F, Liang X, Hartwig FP, Bowden J. The confidence interval method for selecting valid instrumental variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2021;83(4):752–776.

[21] Kreider B, Hill SC. Partially identifying treatment effects with an application to covering the uninsured. Journal of Human Resources 2009;44:409–449.

[22] Imbens G, Manski CF. Confidence intervals for partially identified parameters. Econometrica 2004;72:1845–1857.

[23] Horowitz JL, Manski CF. Identification and estimation of statistical functionals using incomplete data. Journal of Econometrics 2006;132:445–459.

[24] WHO/UNAIDS. Guidelines for measuring national HIV prevalence in population-based surveys. WHO, Geneva; 2005.

[25] Marra G, Radice R, Bärnighausen T, Wood SN, McGovern ME. A simultaneous equation approach to estimating HIV prevalence with nonignorable missing responses. Journal of the American Statistical Association 2017;112(518):484–496.

[26] Jiang Z, Ding P. Using missing types to improve partial identification with application to a study of HIV prevalence in Malawi. The Annals of Applied Statistics 2018;12(3):1831–1852.

[27] Clark SJ, Houle B. Evaluation of Heckman Selection Model Method for Correcting Estimates of HIV Prevalence from Sample Surveys via Realistic Simulation. Cent. Stat. Soc. Sci. Work. Pap. No 120. University of Washington; 2012.

[28] Mishra V, Hong R, Khan S, Gu Y, Liu L. Evaluating HIV estimates from national population-based surveys for bias resulting from non-response. DHS Analytical Studies No. 12. Calverton, Maryland, USA: Macro International; 2008.

[29] Nicoletti C. Poverty analysis with missing data: alternative estimators compared. Empirical Economics 2010;38:1–22.

[30] Kalichman SC, Simbayi LC. HIV testing attitudes, AIDS stigma, and voluntary HIV counselling and testing in a black township in Cape Town, South Africa. Sexually Transmitted Infections 2003;79(6):442–447.

[31] Genberg BL, Hlavka Z, Konda KA, Maman S, Chariyalertsak S, Chingono A, et al. A comparison of HIV/AIDS-related stigma in four countries: Negative attitudes and perceived acts of discrimination towards people living with HIV/AIDS. Social Science & Medicine 2009;68(12):2279–2287.

[32] Kang H, Lee Y, Cai TT, Small DS. Two robust tools for inference about causal effects with invalid instruments. Biometrics 2022;78(1):24–34.

[33] Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 2002;20:518–529.

Table 1: Partial identification of $\mathrm{E}(Y)$ with $L = 5$ instruments and $s = 3$ valid instruments with $\mathrm{E}(Y)$ fixed at 0.15. Results are stratified by average non-response rate $1 - \mathrm{E}(D) = 0.1$ or 0.3; instruments either all strong or a mixture of strong + weak; the last $L - s$ instruments either weakly or strongly violate (4); and mild or moderate selection bias.

Instrument Strength	Non-response rate	Selection bias		Imputation	Worst case bounds	IV bounds
Strong	0.1	Mild	Coverage	0.985	1	0.972
			Lower CI	0.119	0.114	0.123
			Upper CI	0.162	0.262	0.218
			Width	0.043	0.148	0.095
Strong + Weak	0.1	Mild	Coverage	1	1	0.868
			Lower CI	0.125	0.121	0.127
			Upper CI	0.169	0.266	0.248
			Width	0.044	0.145	0.121
Strong	0.3	Mild	Coverage	0.075	1	1
			Lower CI	0.095	0.084	0.108
			Upper CI	0.137	0.434	0.323
			Width	0.042	0.35	0.216
Strong + Weak	0.3	Mild	Coverage	0.078	1	1
			Lower CI	0.088	0.072	0.098
			Upper CI	0.134	0.42	0.287
			Width	0.045	0.349	0.189
Strong	0.1	Moderate	Coverage	0.003	1	1
			Lower CI	0.091	0.088	0.11
			Upper CI	0.129	0.234	0.201
			Width	0.039	0.146	0.092
Strong + Weak	0.1	Moderate	Coverage	0.118	1	1
			Lower CI	0.102	0.098	0.116
			Upper CI	0.142	0.239	0.207
			Width	0.04	0.141	0.091
Strong	0.3	Moderate	Coverage	0	1	1
			Lower CI	0.05	0.046	0.07
			Upper CI	0.081	0.388	0.287
			Width	0.031	0.342	0.218
Strong + Weak	0.3	Moderate	Coverage	0	1	1
			Lower CI	0.065	0.058	0.083
			Upper CI	0.1	0.399	0.313
			Width	0.035	0.341	0.23

Table 2: Summary statistics of 8 scenarios considered in simulation study 2

Interviewer Effects	# invalid IVs	True HIV Prevalence	Observed HIV Prevalence	Consent rate	Contact rate
Weak	3	0.16	0.11	0.75	0.91
Strong	3	0.17	0.13	0.82	0.91
Weak	2	0.19	0.13	0.75	0.91
Strong	2	0.19	0.15	0.81	0.91
Weak	1	0.18	0.12	0.75	0.91
Strong	1	0.18	0.14	0.82	0.91
Weak	0	0.18	0.12	0.75	0.91
Strong	0	0.18	0.14	0.82	0.91

Table 3: Average 95% lower (LCI) and upper (UCI) confidence intervals in 8 scenarios considered in simulation study 2

Interviewer Effects	# Invalid IVs	Imputation		Marra Normal Copula		Marra Best		Worst case bounds		JD bounds		Union bounds	
		LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI	LCI	UCI
Weak	3	0.10	0.12	0.19	0.30	0.19	0.30	0.08	0.34	0.08	0.35	0.09	0.32
Strong	3	0.12	0.14	0.18	0.25	0.18	0.25	0.10	0.30	0.11	0.31	0.12	0.27
Weak	2	0.12	0.13	0.18	0.30	0.18	0.29	0.09	0.36	0.10	0.36	0.11	0.34
Strong	2	0.14	0.15	0.19	0.26	0.19	0.25	0.11	0.32	0.12	0.33	0.14	0.29
Weak	1	0.11	0.13	0.17	0.29	0.17	0.29	0.08	0.35	0.09	0.36	0.09	0.33
Strong	1	0.13	0.15	0.18	0.25	0.18	0.25	0.11	0.31	0.12	0.32	0.12	0.27
Weak	0	0.11	0.13	0.18	0.30	0.18	0.29	0.09	0.35	0.09	0.36	0.09	0.33
Strong	0	0.13	0.15	0.19	0.26	0.19	0.25	0.11	0.31	0.12	0.32	0.12	0.27

Table 4: Width of 95% confidence intervals in 8 scenarios considered in simulation study 2

Interviewer Eff.	# Invalid IVs	Imputation	Marra Normal Copula	Marra Best	Worst case bounds	JD bounds	Union bounds
Weak	3	0.01	0.11	0.10	0.27	0.27	0.23
Strong	3	0.01	0.07	0.07	0.20	0.20	0.15
Weak	2	0.01	0.12	0.12	0.27	0.27	0.23
Strong	2	0.02	0.07	0.07	0.20	0.20	0.15
Weak	1	0.01	0.12	0.12	0.27	0.27	0.23
Strong	1	0.02	0.07	0.07	0.20	0.20	0.15
Weak	0	0.01	0.12	0.12	0.27	0.27	0.23
Strong	0	0.02	0.07	0.07	0.20	0.20	0.15

Table 5: Observed proportions of HIV positive among the tested in 2007 Zambia DHS

	Women		Men	
Age	HIV prevalence	N	HIV prevalence	N
All	0.161	5713	0.123	5163
15-19	0.058	1256	0.035	1109
20-24	0.119	1119	0.053	830
25-29	0.198	1102	0.115	772
30-34	0.258	841	0.174	746
35-39	0.250	588	0.223	594
40-44	0.182	434	0.240	390
45-49	0.122	373	0.183	318

Table 6: Relationship between HIV testing and some possible instrument variables

(a): Women

D		0		1	
Variable	N	Percent	N	Percent	Test
iv.lan	1695		5713		$\chi^2 = 24.706^{***}$
... 0	727	42.9%	2845	49.8%	
... 1	968	57.1%	2868	50.2%	
iv.firstday	1695		5713		$\chi^2 = 22.774^{***}$
... 0	1094	64.5%	3315	58%	
... 1	601	35.5%	2398	42%	
iv.interviewer	1433		5713		$\chi^2 = 28.929^{***}$
... 0	74	5.2%	256	4.5%	
... 1	190	13.3%	807	14.1%	
... 2	695	48.5%	2369	41.5%	
... 3	474	33.1%	2281	39.9%	
iv.mon	1695		5713		$\chi^2 = 82.634^{***}$
... 0	311	18.3%	1688	29.5%	
... 1	1384	81.7%	4025	70.5%	
iv.doa	1695		5713		$\chi^2 = 14.097^{***}$
... 0	688	40.6%	2616	45.8%	
... 1	1007	59.4%	3097	54.2%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

(b): Men

D	0		1		
Variable	N	Percent	N	Percent	Test
iv.lan	1983		5163		$\chi^2 = 160.186^{***}$
... 0	739	37.3%	2789	54%	
... 1	1244	62.7%	2374	46%	
iv.firstday	1983		5163		$\chi^2 = 4.267^{**}$
... 0	1251	63.1%	3118	60.4%	
... 1	732	36.9%	2045	39.6%	
iv.interviewer	1339		5161		$\chi^2 = 13.453^{***}$
... 0	167	12.5%	583	11.3%	
... 1	62	4.6%	373	7.2%	
... 2	493	36.8%	1791	34.7%	
... 3	617	46.1%	2414	46.8%	
iv.mon	1983		5163		$\chi^2 = 200.984^{***}$
... 0	293	14.8%	1621	31.4%	
... 1	1690	85.2%	3542	68.6%	
iv.doa	1983		5163		$\chi^2 = 60.931^{***}$
... 0	650	32.8%	2216	42.9%	
... 1	1333	67.2%	2947	57.1%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

Table 7: 95% confidence intervals for HIV prevalence estimates in 2007 Zambia DHS

(a) Women

			Unadjusted	Marra	Worst case	JD	IV
All		LCI	0.153	0.159	0.118	0.121	0.123
		UCI	0.169	0.179	0.361	0.343	0.339
		Width	0.016	0.02	0.243	0.222	0.216
Age	15-19	LCI	0.038	0.058	0.029	0.029	0.023
		UCI	0.078	0.189	0.305	0.282	0.274
		Width	0.04	0.131	0.276	0.253	0.251
	20-24	LCI	0.099	0.058	0.073	0.076	0.076
		UCI	0.139	0.189	0.36	0.336	0.331
		Width	0.04	0.131	0.287	0.26	0.255
	25-29	LCI	0.181	0.058	0.142	0.144	0.147
		UCI	0.216	0.189	0.392	0.378	0.383
		Width	0.035	0.131	0.25	0.234	0.236
	30-34	LCI	0.237	0.058	0.187	0.191	0.202
		UCI	0.279	0.189	0.44	0.429	0.401
		Width	0.042	0.131	0.253	0.238	0.199
	35-39	LCI	0.223	0.058	0.174	0.179	0.178
		UCI	0.278	0.189	0.443	0.425	0.408
		Width	0.055	0.131	0.269	0.246	0.23
	40-44	LCI	0.151	0.058	0.12	0.121	0.131
		UCI	0.212	0.189	0.391	0.377	0.342
		Width	0.061	0.131	0.271	0.256	0.211
	45-49	LCI	0.09	0.058	0.072	0.073	0.073
		UCI	0.155	0.189	0.333	0.322	0.315
		Width	0.065	0.131	0.261	0.249	0.242

(b) Men

Unadjusted	Marra	Worst case	JD	IV
0.116	0.186	0.084	0.089	0.086
0.13	0.268	0.374	0.328	0.322
0.014	0.082	0.29	0.239	0.236
0.019	0.049	0.014	0.015	0.013
0.05	0.187	0.319	0.271	0.279
0.031	0.138	0.305	0.256	0.266
0.035	0.049	0.025	0.026	0.024
0.072	0.187	0.351	0.302	0.315
0.037	0.138	0.326	0.276	0.291
0.096	0.049	0.067	0.073	0.083
0.134	0.187	0.409	0.353	0.361
0.038	0.138	0.342	0.28	0.278
0.155	0.049	0.111	0.12	0.124
0.192	0.187	0.426	0.382	0.384
0.037	0.138	0.315	0.262	0.26
0.201	0.049	0.147	0.158	0.16
0.244	0.187	0.46	0.417	0.376
0.043	0.138	0.313	0.259	0.216
0.215	0.049	0.166	0.173	0.173
0.264	0.187	0.452	0.421	0.414
0.049	0.138	0.286	0.248	0.241
0.153	0.049	0.115	0.122	0.129
0.212	0.187	0.424	0.386	0.377
0.059	0.138	0.309	0.264	0.248

Supplementary material to HIV estimation using population-based surveys with non-response: a partial identification approach

In this supplement, we provide results for additional materials and clarifications.

S.1 Bounds using instruments

We assume for each individual in the population of interest, an outcome variable Y is measurable. Suppose we are interested in the population mean of Y, $\mathrm{E}(Y)$. In general, we may also be interested in $\mathrm{E}(Y|X)$ for some covariates X, but for brevity, we focus our discussion in the next two sections on estimating $\mathrm{E}(Y)$ since the treatment for the case with covariates is similar. Suppose a random sample of n is drawn from the population, and in this sample, Y is observed only in a subset of the sample. Let D be a binary variable such that $D = 1$ if Y is observed and 0 otherwise. Using the law of iterated expectations, we can write

$$\mathrm{E}(Y) = \mathrm{E}(Y|D=1)\mathrm{P}(D=1) + \mathrm{E}(Y|D=0)\mathrm{P}(D=0). \tag{S.1}$$

The sampling process identifies $\mathrm{E}(Y|D=1)$, $\mathrm{P}(D=1)$ and $\mathrm{P}(D=0) = 1 - \mathrm{P}(D=1)$ but there is no information on $\mathrm{E}(Y|D=0)$ unless we make strong assumptions about the joint distribution of Y and D. Let K_0, K_1 be, respectively, the lower and upper bounds of Y. Furthermore, write $\mu \equiv \mathrm{E}(Y)$, $\mu_{d\cdot} \equiv \mathrm{E}(Y|D=d)$. The worst case partial identification bounds [15] for μ are

$$(\mathrm{LB}, \mathrm{UB}) = (\mu_{1\cdot}\mathrm{P}(D=1) + K_0\mathrm{P}(D=0), \mu_{1\cdot}\mathrm{P}(D=1) + K_1\mathrm{P}(D=0)). \tag{S.2}$$

The worst case bounds (S.2) are guaranteed to identify $\mathrm{E}(Y)$ by construction. However, they are often criticised for being too wide to be informative. The worst case bounds can be improved if additional assumptions are made. Let V be an instrumental variable with discrete values $v \in \mathcal{V}$, such that,

$$\mathrm{P}(D=d|V=v_1) \neq \mathrm{P}(D=d|V=v_2), \tag{S.3}$$

and

$$\mathrm{P}(Y) = \mathrm{P}(Y|V=v_1) = \mathrm{P}(Y|V=v_2), \tag{S.4}$$

for $d = 0, 1$, all values $v_1, v_2 \in \mathcal{V}$ and $v_1 \neq v_2$. Write $\mu_{\cdot v} \equiv \mathrm{E}(Y|V=v)$ and $\mu_{dv} \equiv \mathrm{E}(Y|D=d, V=v)$. Since (S.4) implies $\mathrm{E}(Y|V=v) = \mathrm{E}(Y) = \mu$, it follows that [18], $\forall v \in \mathcal{V}$,

$$\mu_{1v}\mathrm{P}(D=1|V=v) + K_0\mathrm{P}(D=0|V=v) \leq \mu_{\cdot v} \leq \mu_{1v}\mathrm{P}(D=1|V=v) + K_1\mathrm{P}(D=0|V=v).$$

The inequalities imply

$$\begin{aligned}
&\mu \in \bigcap_{v\in\mathcal{V}} [\mu_{1v}\mathrm{P}(D=1|V=v) + K_0\mathrm{P}(D=0|V=v), \mu_{1v}\mathrm{P}(D=1|V=v) + K_1\mathrm{P}(D=0|V=v)] \\
\Rightarrow\quad &\mathrm{LB}_V \equiv \sup_{v\in\mathcal{V}} \{\mu_{1v}\mathrm{P}(D=1|V=v) + K_0\mathrm{P}(D=0|V=v)\} \leq \mu \qquad \text{(S.5)} \\
&\leq \inf_{v\in\mathcal{V}} \{\mu_{1v}\mathrm{P}(D=1|V=v) + K_1\mathrm{P}(D=0|V=v)\} \equiv \mathrm{UB}_V, \qquad \text{(S.6)}
\end{aligned}$$

where $(\mathrm{LB}_V, \mathrm{UB}_V)$ gives a set of IV lower and upper bounds for μ. It is straightforward to see that the IV bounds are guaranteed to lie within the worst case bounds, hence if V is observed for all individuals in the sample, a set of tighter bounds than those given by the worst case bounds can be achieved. Notice that in order for the IV bounds to work, assumptions (S.3) and

(S.4) must both be satisfied. Assumption (S.4) is a necessary condition; violation of (S.4) gives an invalid instrument, which may lead to bounds that fail to identify the quantity of interest. Violation of assumption (S.3) gives a weak instrument [33]. While using a weak instrument does not lead to invalid inferences, the bounds (S.6) become uninformative. To see this last point, suppose (S.4) is satisfied but (S.3) is not, such that $\mathrm{P}(D=d|V=v)=\mathrm{P}(D=d)$ for all $v\in\mathcal{V}$; then the left hand of the inequality (S.6) becomes

$$\begin{aligned}\sup_{v\in\mathcal{V}}\{\mu_{1v}\}\mathrm{P}(D=1)+K_0\mathrm{P}(D=0) &= \sup_{v\in\mathcal{V}}\{\mu_{1\cdot}\}\mathrm{P}(D=1)+K_0\mathrm{P}(D=0)\\ &= \mu_{1\cdot}\mathrm{P}(D=1)+K_0\mathrm{P}(D=0),\end{aligned}$$

which is identical to the lower worst case bound (S.2). Similarly, the right hand side of (S.6) becomes the upper worst case bound. The observed data on D, however, allow us to verify whether an instrument is weak via (S.3).

In practice, more than one instrument is usually used in a particular study [see, e.g., 17, 21]. Suppose l candidate instruments are considered for reducing the width of the worst case bounds. Define $\{V_1,\cdots,V_t\}$ for any arbitrary set of $t\geq 1$ instruments. Suppose there are $t=L>1$ instruments such that $V_l, l=1,\cdots,L$ all satisfy (S.3) and (S.4). Write for V_l, the bounds $(\mathrm{LB}_{V_l},\mathrm{UB}_{V_l})$. Then μ must also lie in the "intersection" of the bounds [19]:

$$(\mathrm{LB}_{\mathrm{IN}},\mathrm{UB}_{\mathrm{IN}})=\underset{\mathcal{V}_l,l=1,\cdots,L}{\cap}(\mathrm{LB}_{V_l},\mathrm{UB}_{V_l})=(\sup_{v\in\mathcal{V}_l}\mathrm{LB}_{V_l},\inf_{v\in\mathcal{V}_l}\mathrm{UB}_{V_l}). \tag{S.7}$$

Even though the IV and intersection bounds provide refinements on the worst case bounds, these refinements are achieved at the expense of having to identify instrumental variables that satisfy assumptions (S.3) together with (S.4). It is well known that valid and informative instruments are difficult to find. More importantly, assumption (S.4) is not verifiable, and hence, in practice, these bounds are anchored on our beliefs that the assumptions are satisfied. If even one of the L instruments is invalid, the bounds would fail to identify $\mathrm{E}(Y)$. This problem, where some of the instruments may be invalid, is well-known in the casual inference literature. Our remedy is to create union bounds:

$$(\mathrm{LB}^{\mathrm{UN}},\mathrm{UB}^{\mathrm{UN}})=\underset{\mathcal{V}_l,l=1,\cdots,L}{\cup}(\mathrm{LB}_{V_l},\mathrm{UB}_{V_l})=(\inf_{v\in\mathcal{V}_l}\mathrm{LB}_{V_l},\sup_{v\in\mathcal{V}_l}\mathrm{UB}_{V_l}). \tag{S.8}$$

It is trivial to see that $(\mathrm{LB}^{\mathrm{UN}},\mathrm{UB}^{\mathrm{UN}})$ identifies $\mathrm{E}(Y)$ as long as at least one of the candidate instruments is valid. To reduce the width of the union bounds, we make the assumption that the true number of valid instruments, s is known to satisfy $s>a\geq 1$ for some known a. Under this assumption, each subset of $(L-a+1)$ instruments must contain at least one valid instrument. Hence, the union bound formed by each subset is guaranteed to identify $\mathrm{E}(Y)$. For any two sets of bounds that both include $E(Y)$, their intersection must be non-empty and also correctly identify $\mathrm{E}(Y)$. We, therefore, propose to find the intersection of all union bounds formed with any $(L-a+1)$ instruments among the L instruments because it will also identify $\mathrm{E}(Y)$ but be no longer than any of these union bounds.

S.2 Confidence intervals

Applying the bounds empirically incurs uncertainty and this uncertainty can be incorporated in the form of confidence intervals. Let $(\mathrm{LB},\mathrm{UB})$ denote a set of generic theoretical lower and upper IV bounds for μ. Let $(\widehat{\mathrm{LB}},\widehat{\mathrm{UB}})$ be any empirical estimate of $(\mathrm{LB},\mathrm{UB})$. A confidence interval should have a high asymptotic probability of containing both $(\mathrm{LB},\mathrm{UB})$ or μ. Here, we focus on finding an approximate $b_0\times 100$ percent for μ. An approximate $b_0\times 100$ percent

confidence interval for (LB, UB) is simply of the form $(\widehat{\text{LB}} - z_{(1-b_0)/2}\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{UB}} + z_{(1-b_0)/2}\widehat{\text{SE}}_{\text{UB}})$, where $z_{(1-b_0)/2}$ is the upper $(1-b_0)/2 \times 100$ percentile of the standard normal distribution, SE represents standard error and $\widehat{\text{SE}}$ its sample analogue. As pointed out by Imbens and Manski [22], this interval would be too wide for μ. In fact, since (LB, UB) is a set of bounds and if we are interested in μ, then it will be nearer to one of $\widehat{\text{LB}}$ or $\widehat{\text{UB}}$ but not both simultaneously. Hence, they suggested the following bounds[2] :

$$(\widehat{\text{LB}} - C_n\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{UB}} + C_n\widehat{\text{SE}}_{\text{UB}})$$

such that C_n is determined by

$$\Phi\left(C_n + \frac{\widehat{\text{UB}} - \widehat{\text{LB}}}{\max(\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{SE}}_{\text{UB}})}\right) - \Phi(-C_n) = b_0,$$

where Φ is the standard normal CDF. For example, if b_0 is 0.95 such that we are interested in approximate 95% confidence intervals, then the value of C_n approaches 1.64 when $\widehat{\text{UB}} - \widehat{\text{LB}}$ is large, and it approaches 1.96 when $\widehat{\text{UB}} - \widehat{\text{LB}}$ is near zero. Since $\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{SE}}_{\text{UB}}$ are extremely difficult to find analytically in all practical cases, following Horowitz and Manski [23], we resort to bootstrapping. We sample with replacement from the data, and we denote a generic bootstrap sample $(d_i^*, v_{1,i}^*, \cdots, v_{L,i}^*, y_i^* \cdot d_i^*)$, where $i = 1, \cdots, n$ is the index for individuals. Using each bootstrap sample, we find $(\widehat{\text{LB}}^*, \widehat{\text{UB}}^*)$ and from B bootstrap samples, we obtain $\widehat{\text{SE}}_{\text{LB}}$ and $\widehat{\text{SE}}_{\text{UB}}$.

S.3 Simulations 1

In this section, we describe additional simulation results. We use a similar setup of the simulation study in the main paper. We assume the response Y is binary. We fix the values of s and L at 3 and 5, respectively. We only consider binary 0-1 instruments; The valid instruments are generated by a multivariate binary distribution, $MVB(\mu_s, \Sigma_{s\times s})$ with $\mu_s = 0.5 \times \mathbb{1}_s$ and $\Sigma_{s\times s} = (\sigma_{jj'}), j, j' = 1, \cdots, s$, where $\sigma_{jj} = 1$ and for $\rho_1 = \sigma_{jj'}, j \neq j'$, we consider two choices of ρ_1: 0 and 0.3. The first choice corresponds to the situation when all valid instruments are mutually independent, while the second choice assumes a correlation of 0.3 between each pair of instruments. We do not believe a high correlation between instruments to be a realistic situation since if two instruments are highly correlated, there is no reason to use both. The invalid instruments are generated independently of the valid instruments using a $MVB(\mu_{L-s}, \Sigma_{(L-s)\times(L-s)})$, with $\mu_{L-s} = 0.5 \times \mathbb{1}_{L-s}$ and $\Sigma_{(L-s)\times(L-s)} = (\sigma_{jj'}), j, j' = 1, \cdots, (L-s)$. We also use the same two choices of 0 and 0.3 for $\rho_2 = \sigma_{jj'}, j \neq j'$ between any two invalid instruments.

We generate Y using a logistic model

$$\text{logit}\text{P}(Y = 1) = b_0 + b_{11}V_1 + \cdots + b_{1L}V_L, \tag{S.9}$$

where the coefficients $b_1 = (b_{11}, \cdots, b_{1L})^T$ give the association between the instruments and Y. A non-zero value of b_{1j} induces an association and therefore renders the instrument invalid. We use two different combinations for b_1: $b_1 = (\overbrace{0, \cdots, 0}^{s}, \overbrace{1, \cdots, 1}^{(L-s)})^T$; and $b_1 = (\overbrace{0, \cdots, 0}^{s}, \overbrace{4, \cdots, 4}^{(L-s)})^T$. For both situations, we assume without loss of generality, the first s instruments are valid while

[2]Our expressions differ from those in equations (6) and (7) of Imbens and Manski [22] by a factor of $\sqrt{n}$ because they use the notation of $\widehat{\sigma}_\cdot/\sqrt{n}$ to denote standard error.

the remaining $L-s$ are invalid. In the former, (4) is weakly violated by the invalid instruments, while the violation of (4) is strong for the latter.

The non-response indicator D is generated using another logistic model

$$\text{logitP}(D=1) = c_0 + c_{11}V_1 + \cdots + c_{1L}V_L + c_Y Y. \tag{S.10}$$

The coefficients $c_1 = (c_{11}, \cdots, c_{1L})^T$ give the association between each instrument and D. We consider two situations: (a) Strong instruments: $c_1 = (5, \cdots, 5)$ and (b) Strong + weak instruments: s coefficients are randomly given a value of 5, and the remaining $L-s$ are given a value of 0.5. The coefficient c_Y is used to model the association of D to the outcome Y. When $c_Y = 0$, then there is no selection bias when conditioned on the observed covariates. We consider two choices of $c_Y = -0.1||c_1||$ and $-0.3||c_1||$, where the symbol $||\cdot||$ stands for the sum of the coefficients $c_{11}, ..., c_{1L}$. We use negative association to reflect that, in practice, we expect those who are HIV positive to be less likely to have an HIV test. These two values for c_Y correspond to weak to moderate associations between Y and D. We use c_0 to calibrate the average non-response rate, $1 - \text{E}(D=1)$, to be 0.1, 0.3, and 0.5 over the simulations.

Since Y is binary, the bounds for Y are $(K_0, K_1) = (0, 1)$. Throughout the study, we use a sample size of $n = 1000$ observations for each simulation run. We use 1000 simulation runs for each combination of parameters and 100 bootstraps to estimate the standard errors of the partial identification bounds.

Tables 1(a)-(c) give the simulation results for $\text{E}(D) = 0.1 - 0.3$, respectively, when Y is weakly negatively associated with D. The corresponding results when Y is moderately associated with D are given in Tables 1(d)-(f).

We consider three different methods for estimating $\text{E}(Y)$: Imputation, partial identification bounds without any assumptions (worst case bounds) and partial identification bounds using instrument variables. For the imputation method, we use all the observed variables in the simulation study, i.e., the instruments. For partial identification, we used the worst case bounds that do not make any assumptions, and also the method proposed in this article.

Each combination of parameters corresponds to four rows of results. The first row shows the proportion of times, out of 1000 simulations, the approximate 95% confidence intervals include $\text{E}(Y)$. The second row gives the lower confidence limits, averaged over 1000 simulations. The third row gives the upper confidence limits, averaged over 1000 simulations. The fourth row gives the average width of the confidence intervals. The results are given in Tables S.1 (a)-(f).

Table S.1: Partial identification of E(Y) with $L = 5$ instruments and $s = 3$ valid instruments; E(Y) fixed at 0.15; $\rho = \rho_1 = \rho_2$ gives correlation between pairs of valid (invalid) instruments; instruments either all strong or a mixture of strong + weak; the last $L - s$ instruments either weakly or strongly violate (4).

(a): Average non-response rate $1 - \mathrm{E}(D) = 0.1$; Y weakly associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0.985	1	0.972
			Lower CI	0.119	0.114	0.123
			Upper CI	0.162	0.262	0.218
			Width	0.043	0.148	0.095
0	Strong	Strong	Coverage	1	1	1
			Lower CI	0.125	0.124	0.126
			Upper CI	0.175	0.28	0.224
			Width	0.049	0.156	0.098
0	Strong + Weak	Weak	Coverage	1	1	0.868
			Lower CI	0.125	0.121	0.127
			Upper CI	0.169	0.266	0.248
			Width	0.044	0.145	0.121
0	Strong + Weak	Strong	Coverage	1	1	0.999
			Lower CI	0.124	0.122	0.123
			Upper CI	0.169	0.27	0.265
			Width	0.045	0.148	0.142
0.3	Strong	Weak	Coverage	0.988	1	0.893
			Lower CI	0.115	0.111	0.114
			Upper CI	0.168	0.269	0.226
			Width	0.053	0.157	0.111
0.3	Strong	Strong	Coverage	1	1	0.998
			Lower CI	0.081	0.081	0.08
			Upper CI	0.22	0.356	0.25
			Width	0.139	0.275	0.17
0.3	Strong + Weak	Weak	Coverage	1	1	0.886
			Lower CI	0.122	0.119	0.124
			Upper CI	0.168	0.267	0.244
			Width	0.046	0.149	0.12
0.3	Strong + Weak	Strong	Coverage	1	1	1
			Lower CI	0.124	0.124	0.125
			Upper CI	0.173	0.281	0.252
			Width	0.05	0.157	0.127

(b): Average non-response rate $1 - \mathrm{E}(D) = 0.3$; Y weakly associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0.075	1	1
			Lower CI	0.095	0.084	0.108
			Upper CI	0.137	0.434	0.323
			Width	0.042	0.35	0.216
0	Strong	Strong	Coverage	0.959	1	0.991
			Lower CI	0.114	0.105	0.123
			Upper CI	0.157	0.46	0.341
			Width	0.043	0.355	0.218
0	Strong + Weak	Weak	Coverage	0.078	1	1
			Lower CI	0.088	0.072	0.098
			Upper CI	0.134	0.42	0.287
			Width	0.045	0.349	0.189
0	Strong + Weak	Strong	Coverage	1	1	0.982
			Lower CI	0.128	0.126	0.134
			Upper CI	0.172	0.478	0.447
			Width	0.044	0.353	0.312
0.3	Strong	Weak	Coverage	0.166	1	0.999
			Lower CI	0.098	0.087	0.117
			Upper CI	0.142	0.438	0.275
			Width	0.044	0.351	0.158
0.3	Strong	Strong	Coverage	0.825	1	0.974
			Lower CI	0.113	0.105	0.128
			Upper CI	0.156	0.449	0.275
			Width	0.043	0.345	0.147
0.3	Strong + Weak	Weak	Coverage	0.437	1	1
			Lower CI	0.098	0.078	0.114
			Upper CI	0.149	0.424	0.223
			Width	0.051	0.346	0.11
0.3	Strong + Weak	Strong	Coverage	0.71	1	0.997
			Lower CI	0.101	0.078	0.114
			Upper CI	0.153	0.424	0.222
			Width	0.052	0.346	0.107

(c): Average non-response rate $1 - \mathrm{E}(D) = 0.5$; Y weakly associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0.072	1	1
			Lower CI	0.077	0.061	0.085
			Upper CI	0.127	0.607	0.472
			Width	0.049	0.546	0.387
0	Strong	Strong	Coverage	0.624	1	1
			Lower CI	0.105	0.086	0.111
			Upper CI	0.151	0.633	0.498
			Width	0.046	0.547	0.386
0	Strong + Weak	Weak	Coverage	0.314	1	1
			Lower CI	0.083	0.061	0.086
			Upper CI	0.142	0.606	0.454
			Width	0.059	0.545	0.368
0	Strong + Weak	Strong	Coverage	0.723	1	1
			Lower CI	0.106	0.079	0.111
			Upper CI	0.157	0.624	0.514
			Width	0.051	0.545	0.403
0.3	Strong	Weak	Coverage	0.137	1	1
			Lower CI	0.081	0.064	0.098
			Upper CI	0.133	0.607	0.409
			Width	0.052	0.543	0.311
0.3	Strong	Strong	Coverage	0.459	1	1
			Lower CI	0.1	0.079	0.116
			Upper CI	0.149	0.629	0.429
			Width	0.049	0.55	0.313
0.3	Strong + Weak	Weak	Coverage	0.368	1	1
			Lower CI	0.082	0.061	0.096
			Upper CI	0.144	0.609	0.49
			Width	0.062	0.548	0.394
0.3	Strong + Weak	Strong	Coverage	0.669	1	0.999
			Lower CI	0.098	0.072	0.112
			Upper CI	0.154	0.618	0.502
			Width	0.056	0.546	0.39

(d): Average non-response rate $1 - \mathrm{E}(D) = 0.1$; Y moderately associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0.003	1	1
			Lower CI	0.091	0.088	0.11
			Upper CI	0.129	0.234	0.201
			Width	0.039	0.146	0.092
0	Strong	Strong	Coverage	0.88	1	1
			Lower CI	0.114	0.108	0.113
			Upper CI	0.156	0.254	0.21
			Width	0.042	0.146	0.097
0	Strong + Weak	Weak	Coverage	0.118	1	1
			Lower CI	0.102	0.098	0.116
			Upper CI	0.142	0.239	0.207
			Width	0.04	0.141	0.091
0	Strong + Weak	Strong	Coverage	0.403	1	1
			Lower CI	0.108	0.095	0.105
			Upper CI	0.151	0.237	0.194
			Width	0.043	0.143	0.089
0.3	Strong	Weak	Coverage	0.161	1	0.993
			Lower CI	0.102	0.099	0.115
			Upper CI	0.142	0.244	0.194
			Width	0.04	0.146	0.079
0.3	Strong	Strong	Coverage	0.974	1	0.999
			Lower CI	0.098	0.094	0.098
			Upper CI	0.179	0.29	0.213
			Width	0.081	0.195	0.115
0.3	Strong + Weak	Weak	Coverage	0.662	1	0.982
			Lower CI	0.11	0.108	0.112
			Upper CI	0.151	0.255	0.231
			Width	0.041	0.147	0.119
0.3	Strong + Weak	Strong	Coverage	0.929	1	1
			Lower CI	0.109	0.104	0.108
			Upper CI	0.161	0.266	0.224
			Width	0.052	0.162	0.116

(e): Average non-response rate $1 - \mathrm{E}(D) = 0.3$; Y moderately associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.05	0.046	0.07
			Upper CI	0.081	0.388	0.287
			Width	0.031	0.342	0.218
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.079	0.069	0.098
			Upper CI	0.117	0.413	0.302
			Width	0.038	0.344	0.204
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.065	0.058	0.083
			Upper CI	0.1	0.399	0.313
			Width	0.035	0.341	0.23
0	Strong + Weak	Strong	Coverage	0.006	1	1
			Lower CI	0.088	0.074	0.106
			Upper CI	0.129	0.423	0.326
			Width	0.041	0.349	0.22
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.067	0.062	0.097
			Upper CI	0.102	0.404	0.252
			Width	0.035	0.342	0.155
0.3	Strong	Strong	Coverage	0	1	0.999
			Lower CI	0.085	0.077	0.115
			Upper CI	0.123	0.424	0.257
			Width	0.038	0.347	0.143
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.063	0.057	0.091
			Upper CI	0.099	0.407	0.212
			Width	0.036	0.35	0.122
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.069	0.059	0.099
			Upper CI	0.108	0.404	0.21
			Width	0.039	0.345	0.111

(f): Average non-response rate $1 - \mathrm{E}(D) = 0.5$; Y moderately associated with D

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.029	0.026	0.044
			Upper CI	0.054	0.57	0.44
			Width	0.025	0.544	0.396
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.054	0.045	0.072
			Upper CI	0.087	0.588	0.455
			Width	0.033	0.543	0.383
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.041	0.037	0.061
			Upper CI	0.071	0.582	0.448
			Width	0.03	0.545	0.386
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.052	0.04	0.069
			Upper CI	0.089	0.584	0.481
			Width	0.038	0.543	0.412
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.04	0.037	0.065
			Upper CI	0.069	0.575	0.382
			Width	0.029	0.539	0.316
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.056	0.05	0.088
			Upper CI	0.09	0.595	0.394
			Width	0.034	0.545	0.306
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.037	0.034	0.06
			Upper CI	0.067	0.583	0.464
			Width	0.029	0.548	0.404
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.054	0.046	0.085
			Upper CI	0.09	0.587	0.472
			Width	0.036	0.541	0.387

S.3.1 Simulations 1 Sensitivity analysis

In practice, selecting the value of s may be challenging. To address this, we carried out a sensitivity analysis that presents bounds for different values of s as a sensitivity parameter. The analysis uses the same setup as S.3 except we allow s to range from 0, i.e., no valid instruments among L candidate instruments, to L, i.e., all candidate instruments are valid. For the value of a, the number of instruments that we believed is valid, we considered two values, 3 and 1. For example, if $a = 3$ whereas in fact $s < 3$, then our assumption of $s \geq a \geq 1$ does not hold. For brevity, we only show results for the case where the non-response rate is $1 - \mathrm{E}(D) = 0.3$ and Y and D are moderately correlated. The results for the other cases can be obtained upon request from the corresponding author.

We first present results for $a = 3$ in Table S.2(a)-(f), corresponding to $s = 0$ to 5. For these simulations, we know (a)-(c) correspond to situations where our assumption of $s \geq a \geq 1$ is violated. Hence we don't expect the method proposed in this paper to work well. However, when we observe the results in (a)-(c), we notice that even though the proposed method doesn't work well in cases where the valid instrument assumption (4) is strongly violated, when (4) is only weakly violated, its performance still holds up. The remaining tables (d)-(f) correspond to cases where our assumption of $s \geq a \geq 1$ is valid, the method works well as expected.

Next, we turn to the results when $a = 1$ (Table S.3(a)-(f)). In this case, $s \geq a \geq 1$ is violated only when $s = 0$ (Table S.3(a)). As expected, the results for the proposed method is not satisfactory here. In the remaining tables (b)-(f), the assumption $s \geq a \geq 1$ is satisfied; we can observe that the proposed method properly includes the unknown HIV rate.

Comparing Tables S.2 and S.3, we notice that for each value of s and each scenario, the corresponding the CI assuming $a = 3$ (Table S.2) is narrower than that for $a = 1$ (Table S.3). This is to be expected since assuming $a = 3$, we are taking the intersection of a larger subset of instruments each time, and that reduces the width. However, the reduction comes at a price if the assumed number of valid instruments a exceeds the actual number s when the shorter bounds fail to capture the unknown HIV rate. Hence, we need to strike a balance. Obviously, in practice, we must have certain confidence in the validity of the candidate instruments before they should be included into consideration for creating the bounds.

Table S.2: Sensitivity analysis of partial identification methods. The number of instruments L fixed at 5 and the number of valid instruments s ranging from 0 to 5 and the number of instruments assumed to be valid $a = 3$; $\mathrm{E}(Y)$ fixed at 0.15; $\rho = \rho_1 = \rho_2$ gives correlation between pairs of valid (invalid) instruments; instruments either all strong or a mixture of strong + weak; the last $L - s$ instruments either weakly or strongly violate (4). Average non-response rate $1 - \mathrm{E}(D) = 0.3$; Y moderately associated with D

(a): Number of valid instruments $s = 0$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.068	0.066	0.108
			Upper CI	0.103	0.417	0.342
			Width	0.035	0.351	0.234
0	Strong	Strong	Coverage	0.998	1	0
			Lower CI	0.117	0.115	0.196
			Upper CI	0.16	0.465	0.424
			Width	0.043	0.35	0.228
0	Strong + Weak	Weak	Coverage	0.6	1	1
			Lower CI	0.104	0.08	0.114
			Upper CI	0.15	0.431	0.422
			Width	0.046	0.35	0.308
0	Strong + Weak	Strong	Coverage	1	1	0.392
			Lower CI	0.126	0.096	0.161
			Upper CI	0.172	0.446	0.411
			Width	0.047	0.35	0.251
0.3	Strong	Weak	Coverage	0.95	1	0
			Lower CI	0.114	0.113	0.202
			Upper CI	0.157	0.462	0.34
			Width	0.043	0.349	0.138
0.3	Strong	Strong	Coverage	1	1	0
			Lower CI	0.128	0.128	0.255
			Upper CI	0.172	0.481	0.392
			Width	0.044	0.353	0.138
0.3	Strong + Weak	Weak	Coverage	0.934	1	0.154
			Lower CI	0.11	0.091	0.158
			Upper CI	0.155	0.441	0.43
			Width	0.045	0.35	0.272
0.3	Strong + Weak	Strong	Coverage	1	1	0
			Lower CI	0.125	0.103	0.205
			Upper CI	0.17	0.453	0.43
			Width	0.045	0.349	0.225

(b): Number of valid instruments $s = 1$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.065	0.061	0.1
			Upper CI	0.099	0.409	0.336
			Width	0.035	0.348	0.236
0	Strong	Strong	Coverage	0.006	1	0.162
			Lower CI	0.096	0.092	0.162
			Upper CI	0.137	0.443	0.41
			Width	0.041	0.351	0.248
0	Strong + Weak	Weak	Coverage	0.02	1	1
			Lower CI	0.092	0.085	0.116
			Upper CI	0.134	0.436	0.415
			Width	0.042	0.351	0.299
0	Strong + Weak	Strong	Coverage	0.888	1	0.002
			Lower CI	0.109	0.103	0.172
			Upper CI	0.152	0.447	0.39
			Width	0.043	0.344	0.218
0.3	Strong	Weak	Coverage	0.178	1	0
			Lower CI	0.105	0.103	0.184
			Upper CI	0.147	0.442	0.332
			Width	0.041	0.339	0.148
0.3	Strong	Strong	Coverage	1	1	0
			Lower CI	0.128	0.128	0.255
			Upper CI	0.173	0.471	0.394
			Width	0.044	0.343	0.139
0.3	Strong + Weak	Weak	Coverage	0.432	1	0.016
			Lower CI	0.104	0.1	0.168
			Upper CI	0.146	0.45	0.419
			Width	0.042	0.35	0.251
0.3	Strong + Weak	Strong	Coverage	1	1	0
			Lower CI	0.126	0.123	0.245
			Upper CI	0.17	0.474	0.454
			Width	0.044	0.351	0.209

(c): Number of valid instruments $s = 2$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.057	0.053	0.086
			Upper CI	0.09	0.398	0.322
			Width	0.033	0.345	0.237
0	Strong	Strong	Coverage	0.013	1	0.359
			Lower CI	0.094	0.081	0.152
			Upper CI	0.134	0.428	0.406
			Width	0.041	0.347	0.254
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.071	0.064	0.09
			Upper CI	0.11	0.411	0.374
			Width	0.039	0.348	0.284
0	Strong + Weak	Strong	Coverage	0.204	1	0.798
			Lower CI	0.101	0.079	0.141
			Upper CI	0.144	0.433	0.336
			Width	0.043	0.354	0.195
0.3	Strong	Weak	Coverage	0	1	0.972
			Lower CI	0.079	0.074	0.131
			Upper CI	0.117	0.427	0.326
			Width	0.038	0.352	0.196
0.3	Strong	Strong	Coverage	0.746	1	0
			Lower CI	0.109	0.099	0.194
			Upper CI	0.152	0.442	0.396
			Width	0.043	0.343	0.202
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.079	0.071	0.115
			Upper CI	0.118	0.419	0.357
			Width	0.039	0.348	0.242
0.3	Strong + Weak	Strong	Coverage	1	1	0
			Lower CI	0.126	0.126	0.248
			Upper CI	0.17	0.481	0.462
			Width	0.044	0.355	0.214

(d): Number of valid instruments $s = 3$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.049	0.045	0.068
			Upper CI	0.08	0.39	0.287
			Width	0.031	0.344	0.219
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.078	0.068	0.096
			Upper CI	0.118	0.411	0.301
			Width	0.039	0.343	0.204
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.062	0.056	0.08
			Upper CI	0.098	0.409	0.318
			Width	0.036	0.353	0.238
0	Strong + Weak	Strong	Coverage	0.004	1	1
			Lower CI	0.09	0.076	0.108
			Upper CI	0.132	0.42	0.319
			Width	0.042	0.344	0.212
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.067	0.063	0.097
			Upper CI	0.103	0.403	0.251
			Width	0.036	0.34	0.154
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.081	0.073	0.11
			Upper CI	0.12	0.424	0.256
			Width	0.039	0.35	0.146
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.068	0.062	0.098
			Upper CI	0.105	0.408	0.315
			Width	0.037	0.346	0.217
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.082	0.073	0.113
			Upper CI	0.122	0.421	0.327
			Width	0.04	0.348	0.214

(e): Number of valid instruments $s = 4$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.04	0.037	0.056
			Upper CI	0.07	0.384	0.275
			Width	0.029	0.347	0.219
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.052	0.047	0.068
			Upper CI	0.085	0.389	0.277
			Width	0.033	0.342	0.208
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.053	0.049	0.071
			Upper CI	0.086	0.39	0.278
			Width	0.033	0.341	0.207
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.068	0.062	0.088
			Upper CI	0.104	0.404	0.288
			Width	0.036	0.342	0.2
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.062	0.057	0.09
			Upper CI	0.098	0.404	0.225
			Width	0.035	0.347	0.135
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.074	0.067	0.104
			Upper CI	0.112	0.42	0.235
			Width	0.038	0.353	0.132
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.06	0.055	0.088
			Upper CI	0.095	0.405	0.237
			Width	0.035	0.349	0.149
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.07	0.064	0.099
			Upper CI	0.108	0.413	0.239
			Width	0.037	0.349	0.141

(f): Number of valid instruments $s = 5$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.034	0.031	0.046
			Upper CI	0.061	0.37	0.259
			Width	0.027	0.339	0.213
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.033	0.031	0.045
			Upper CI	0.06	0.372	0.261
			Width	0.027	0.342	0.216
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.036	0.033	0.048
			Upper CI	0.063	0.365	0.255
			Width	0.028	0.332	0.207
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.034	0.031	0.046
			Upper CI	0.061	0.37	0.259
			Width	0.027	0.339	0.213
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.063	0.057	0.09
			Upper CI	0.099	0.399	0.216
			Width	0.036	0.342	0.126
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.062	0.057	0.09
			Upper CI	0.098	0.404	0.218
			Width	0.036	0.347	0.128
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.062	0.057	0.09
			Upper CI	0.098	0.397	0.215
			Width	0.036	0.34	0.125
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.062	0.056	0.089
			Upper CI	0.098	0.405	0.219
			Width	0.036	0.349	0.129

Table S.3: Sensitivity analysis of partial identification methods. The number of instruments L fixed at 5 and the number of valid instruments s ranging from 0 to 5 and the number of instruments assumed to be valid $a = 1$; $\mathrm{E}(Y)$ fixed at 0.15; $\rho = \rho_1 = \rho_2$ gives correlation between pairs of valid (invalid) instruments; instruments either all strong or a mixture of strong + weak; the last $L - s$ instruments either weakly or strongly violate (4). Average non-response rate $1 - \mathrm{E}(D) = 0.3$; Y moderately associated with D

(a): Number of valid instruments $s = 0$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.069	0.067	0.1
			Upper CI	0.105	0.418	0.365
			Width	0.036	0.351	0.265
0	Strong	Strong	Coverage	0.993	1	0.001
			Lower CI	0.117	0.114	0.183
			Upper CI	0.16	0.466	0.444
			Width	0.043	0.352	0.261
0	Strong + Weak	Weak	Coverage	0.65	1	1
			Lower CI	0.103	0.08	0.103
			Upper CI	0.15	0.429	0.427
			Width	0.046	0.349	0.324
0	Strong + Weak	Strong	Coverage	1	1	0.725
			Lower CI	0.123	0.094	0.147
			Upper CI	0.169	0.442	0.426
			Width	0.046	0.348	0.279
0.3	Strong	Weak	Coverage	0.945	1	0
			Lower CI	0.113	0.112	0.191
			Upper CI	0.155	0.462	0.354
			Width	0.042	0.351	0.164
0.3	Strong	Strong	Coverage	1	1	0
			Lower CI	0.132	0.132	0.252
			Upper CI	0.176	0.481	0.413
			Width	0.045	0.349	0.161
0.3	Strong + Weak	Weak	Coverage	0.929	1	0.376
			Lower CI	0.113	0.093	0.154
			Upper CI	0.159	0.445	0.442
			Width	0.046	0.352	0.288
0.3	Strong + Weak	Strong	Coverage	1	1	0.001
			Lower CI	0.127	0.105	0.201
			Upper CI	0.172	0.455	0.443
			Width	0.045	0.35	0.242

(b): Number of valid instruments $s = 1$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.063	0.06	0.079
			Upper CI	0.098	0.409	0.357
			Width	0.034	0.349	0.278
0	Strong	Strong	Coverage	0.001	1	1
			Lower CI	0.093	0.09	0.112
			Upper CI	0.133	0.443	0.427
			Width	0.04	0.354	0.315
0	Strong + Weak	Weak	Coverage	0.02	1	1
			Lower CI	0.092	0.085	0.088
			Upper CI	0.133	0.431	0.425
			Width	0.042	0.346	0.337
0	Strong + Weak	Strong	Coverage	0.07	1	1
			Lower CI	0.096	0.07	0.111
			Upper CI	0.14	0.416	0.383
			Width	0.044	0.346	0.272
0.3	Strong	Weak	Coverage	0.098	1	1
			Lower CI	0.101	0.099	0.106
			Upper CI	0.142	0.449	0.36
			Width	0.041	0.351	0.254
0.3	Strong	Strong	Coverage	1	1	0.998
			Lower CI	0.127	0.126	0.129
			Upper CI	0.171	0.474	0.41
			Width	0.044	0.348	0.281
0.3	Strong + Weak	Weak	Coverage	0.404	1	1
			Lower CI	0.107	0.103	0.105
			Upper CI	0.15	0.45	0.439
			Width	0.043	0.347	0.334
0.3	Strong + Weak	Strong	Coverage	1	1	0.998
			Lower CI	0.124	0.121	0.123
			Upper CI	0.168	0.472	0.466
			Width	0.044	0.351	0.343

(c): Number of valid instruments $s = 2$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.055	0.051	0.066
			Upper CI	0.087	0.401	0.349
			Width	0.033	0.35	0.283
0	Strong	Strong	Coverage	0.009	1	1
			Lower CI	0.095	0.083	0.099
			Upper CI	0.137	0.433	0.428
			Width	0.041	0.35	0.328
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.07	0.063	0.065
			Upper CI	0.108	0.413	0.404
			Width	0.038	0.35	0.339
0	Strong + Weak	Strong	Coverage	0.683	1	1
			Lower CI	0.113	0.11	0.112
			Upper CI	0.155	0.457	0.451
			Width	0.042	0.347	0.339
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.082	0.078	0.094
			Upper CI	0.12	0.419	0.344
			Width	0.038	0.342	0.25
0.3	Strong	Strong	Coverage	0.66	1	1
			Lower CI	0.109	0.098	0.115
			Upper CI	0.152	0.446	0.419
			Width	0.043	0.347	0.304
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.08	0.072	0.081
			Upper CI	0.119	0.417	0.413
			Width	0.04	0.344	0.332
0.3	Strong + Weak	Strong	Coverage	0.129	1	1
			Lower CI	0.098	0.084	0.093
			Upper CI	0.142	0.439	0.401
			Width	0.043	0.355	0.308

(d): Number of valid instruments $s = 3$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.04	0.037	0.048
			Upper CI	0.07	0.384	0.328
			Width	0.029	0.347	0.28
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.051	0.046	0.059
			Upper CI	0.084	0.394	0.388
			Width	0.033	0.348	0.33
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.052	0.048	0.05
			Upper CI	0.084	0.394	0.385
			Width	0.032	0.345	0.335
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.067	0.061	0.063
			Upper CI	0.103	0.409	0.402
			Width	0.036	0.348	0.338
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.062	0.057	0.082
			Upper CI	0.098	0.404	0.374
			Width	0.035	0.347	0.292
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.072	0.066	0.092
			Upper CI	0.11	0.414	0.401
			Width	0.038	0.348	0.309
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.062	0.056	0.077
			Upper CI	0.098	0.4	0.375
			Width	0.036	0.344	0.299
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.072	0.066	0.084
			Upper CI	0.11	0.418	0.412
			Width	0.038	0.352	0.328

(e): Number of valid instruments $s = 4$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.04	0.037	0.048
			Upper CI	0.07	0.384	0.328
			Width	0.029	0.347	0.28
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.051	0.046	0.059
			Upper CI	0.084	0.394	0.388
			Width	0.033	0.348	0.33
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.052	0.048	0.05
			Upper CI	0.084	0.394	0.385
			Width	0.032	0.345	0.335
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.067	0.061	0.063
			Upper CI	0.103	0.409	0.402
			Width	0.036	0.348	0.338
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.062	0.057	0.082
			Upper CI	0.098	0.404	0.374
			Width	0.035	0.347	0.292
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.072	0.066	0.092
			Upper CI	0.11	0.414	0.401
			Width	0.038	0.348	0.309
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.062	0.056	0.077
			Upper CI	0.098	0.4	0.375
			Width	0.036	0.344	0.299
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.072	0.066	0.084
			Upper CI	0.11	0.418	0.412
			Width	0.038	0.352	0.328

(f): Number of valid instruments $s = 5$

ρ	Instrument Strength	Violation of (4)		Imputation	Worst case bounds	IV bounds
0	Strong	Weak	Coverage	0	1	1
			Lower CI	0.034	0.031	0.04
			Upper CI	0.061	0.37	0.28
			Width	0.028	0.339	0.24
0	Strong	Strong	Coverage	0	1	1
			Lower CI	0.033	0.031	0.039
			Upper CI	0.06	0.372	0.282
			Width	0.027	0.342	0.243
0	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.034	0.032	0.041
			Upper CI	0.062	0.367	0.279
			Width	0.028	0.335	0.238
0	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.034	0.031	0.04
			Upper CI	0.061	0.371	0.282
			Width	0.027	0.341	0.242
0.3	Strong	Weak	Coverage	0	1	1
			Lower CI	0.062	0.057	0.082
			Upper CI	0.098	0.398	0.232
			Width	0.036	0.34	0.15
0.3	Strong	Strong	Coverage	0	1	1
			Lower CI	0.061	0.056	0.081
			Upper CI	0.097	0.404	0.234
			Width	0.036	0.347	0.153
0.3	Strong + Weak	Weak	Coverage	0	1	1
			Lower CI	0.063	0.057	0.082
			Upper CI	0.099	0.403	0.235
			Width	0.036	0.346	0.152
0.3	Strong + Weak	Strong	Coverage	0	1	1
			Lower CI	0.062	0.056	0.082
			Upper CI	0.098	0.412	0.239
			Width	0.036	0.355	0.158

S.4 Explanation of the set up of Simulations 2

In this section, we report details of a second set of simulations that allows comparison between the proposed method and methods from Marra et al. [25] and Jiang and Ding [26]. The simulations follow the approach of Clark and Houle [27] by using data with a structure similar to a real DHS survey. We use the 2007 Zambia DHS men sub-sample as the basis of our simulation setup.

The relevant individuals in the survey are those men who were eligible for individual surveys. In the 2007 Zambia DHS, individuals were first approached for the individual surveys. Those who were contactable and present at the individual surveys were then asked to participate in HIV testing. Therefore, eligible individuals can be classified into one of three groups: (a) those who were absent for the individual surveys and not tested, (b) those who participated in the individual surveys but refused to be tested, (c) those who participated in the individual surveys, agreed to be tested and with valid test results. For those in groups (a) and (b), their HIV test results are absent. For the simulation setup, HIV results are generated through a three-stage process: (1) contact for individual surveys, (2) consent to HIV test, and (3) test results among the tested. In the simulations, we mimic this three-stage process. First, we fix a sample size of $n = 7000$. Then, we simulate observations with a composition of age, residence (rural vs. urban) and region (Central, Copperbelt, Eastern, Luapula, Lusaka, Northern, Northwestern, Southern, and Western) similar to that of the original survey. To match residence-region composition, we first identify the 319 survey clusters from the survey. We then sample with replacements from these clusters to reach $n = 7000$. We then generate random age identifiers for each of these 7000 observations to match the distribution in the survey.

We also need to create the instrumental variables (IVs) for the methods that require IVs. Following Marra et al. [25], we randomly generate an interviewer IV with 30 interviewers. We additionally generate three binary IVs: V_2 V_3, V_4, each with a prevalence of 0.5 in the samples.

To create the three sub-samples (a)-(c), we make use of three different models. First, we create a contact equation using

$$c = a_0 + a_{11}\text{residence} + a_{12}\text{age} + a_{13}\text{region} + u_c, \tag{S.11}$$

where u_c is a standard normal random variable. We define a binary contact variable $C = 1$ if $c > 0$ and $C = 0$ otherwise. The values of the parameters a_{11}, a_{12}, a_{13} are obtained from fitting a similar model to the actual 2007 Zambia DHS data, and a_0 is set such that the simulated sample has non-contact rate of about 10%.

We use a set of two equations: A selection equation and an outcome equation, on the subset of observations with $C = 1$. The selection equation is defined as

$$d = b_0 + b_{11}\text{residence} + b_{12}\text{age} + b_{13}\text{region} + b_{14}\text{interviewer} + b_2V_2 + b_3V_3 + b_4V_4 + u_d, \tag{S.12}$$

and the outcome equation is defined as

$$y = c_0 + c_{11}\text{residence} + c_{12}\text{age} + c_{13}\text{region} + c_{14}\text{interviewer} + c_2V_2 + c_3V_3 + c_4V_4 + u_y + u_r, \tag{S.13}$$

where (u_d, u_y) are generated using a $MVN(0, \Sigma)$ distribution. The quantity u_r is used to model spatial correlations in HIV rates. We obtain GPS coordinates of each of the 319 clusters and then use generate u_r using an exponential variogram with a sill of 1, range of 1 based on these coordinates. Since the coordinates are in degrees, one unit is equivalent to about 111 km and so our model assumes there is no correlation in HIV rates between two locations beyond 111 km apart in longitude or latitude. The parameters $b_{11}, \cdots, b_{14}$ and $c_{11}, \cdots, c14$ are obtained

by fitting a Heckman selection model to the DHS men data using these equations (omitting V_2, V_3, V_4 since these are hypothetical and not observed in the DHS data). The coefficients $b_2 - b_4$ and $c_2 - c_4$ are adjusted in the simulations to reflect the effects and validity of V_2, V_3, and V_4. The parameters b_0 and c_0 are set to create a refusal rate of about 25% and HIV prevalence of about 20% in the simulated data.

The values b_{14}, b_2, b_3 and b_4 represent IV strength. In the simulations, we let $b_{14} = b_2 = b_3 = b_4$ and we use two different values 0.25 (Weak) and 0.5 (Strong). Throughout the simulations we fix V_4 as a valid IV (ie., $c_4 = 0$). We give c_2, c_3 or c_{14} a value of 0.25 to render the corresponding IV invalid. We consider four different cases in terms of the number of invalid instruments:: (1) three (V_2, V_3, and Interviewer IV) are invalid ($c_2 = c_3 = c_{14} = 0.25$), (2) two ($V_2$ and V_3) are invalid ($c_2 = c_3 = 0.25; c_{14} = 0$), (3) one ($V_2$) is invalid ($c_2 = 0.25; c_3 = c_{14} = 0$), and (4) none is invalid ($c_2 = c_3 = c_{14} = 0$). In (S.12), the Interviewer IV represents the interviewers' differential persuasiveness in eliciting acceptance to HIV test. We generate random $N(0, 1)$ to represent their effectiveness. In (S.13), we generate a separate set of $N(0, 1)$ for each of these interviewers and when $c_{14} \neq 0$, they represent different degrees of IV violation.

For the method of Marra et al. [25], we use only interviewer as the IV. The procedure is implemented via three equations, a selection equation with age, rural, region as predictors, an outcome equation with the same variables as predictors and a third equation that models the copula between selection and outcome (see Heckman [13]) using region only as predictor. In the simulations, we try a selection of representative copulas: Normal, Frank, Clayton rotated 90 degrees and Clayton rotated 270 degrees and then choose the best among them based on AIC. The effect of region is modelled using a Markov random field smoother (see Marra et al. [25]).

The method proposed here always uses all $L = 4$ IVs and assumes at least $a = 2$ of them are valid. Under scenario (1), the assumption is violated. Therefore, the four cases of values allow us to test the sensitivity of the method to the choice s.

For the Jiang and Ding [26] method, all observations with $C = 0$ (non-contact) are assumed to have the same prevalence as those tested.

Results of the simulations are given in Tables 2, 3, and 4 in the main text.

S.5 HIV prevalence using 2007 Zambia DHS

We use the 2007 Zambia DHS data to study non-response adjustment using partial identification bounds. We compare the results to conventional non-response adjustment using imputation. For partial identification, we consider worst case bounds without making any assumptions and also instrumental variable bounds. For the instrumental variable bounds, we use six candidate instrumental variables: iv.lan (whether the language used in the questionnaire or interview is the same as the respondent's language, yes *vs.* no), iv.firstday (whether the interview was conducted on the first day of the interviews, yes *vs.* no), iv.interviewer (number of interviews the interviewer has performed, < 50, $50 - 100$, $100 - 200$, > 200), iv.mon (whether the interview was carried out during a month of harvest or planting, yes *vs.* no), iv.doa (whether the respondent has known someone who has died of AIDS yes *vs.* no).

The standard non-response adjustment is an imputation analysis on those who are not tested to adjust for potential biases [28]. The individuals in the survey can be classified into one of three groups: (a) those who participated in the household and individual surveys and were tested (b) those who participated in the household and individual surveys but were not tested, and (c) those who only participated in the household surveys. For those in groups (b) and (c), their HIV test results are absent.

For individuals in groups (b) and (c), their probability of HIV is predicted based on multivariate models using data from those who were tested. A logistic regression model is used to calculate HIV robability separately for groups (b) and (c). For group (b), the variables used in the model include the following household survey variables: age, education, wealth quintile, residence, and geographic region, as well as the following variables from the individual survey: marital union, current work status, media exposure, eligion, STI or STI symptoms in past 12 months, cigarette smoking/tobacco use, age at first sex, number of sex partners in past 12 months, higher-risk sex in past 12 months, condom use at last sex in past 12 months, and willingness to care for a family member with AIDS. Prediction for group (c) uses only the household variables. The models are used to impute HIV statuses for individuals in groups (b) and (c), and the results are combined with those in group (a) to form adjusted HIV prevalence estimates for the population.

We require instruments for the partial identification bounds method proposed in this paper. We consider five candidate instruments: iv.lan (whether the language used in the questionnaire or interview is the same as the respondent's language, yes *vs.* no), iv.firstday (whether the interview was conducted on the first day of the interviews, yes *vs.* no), iv.interviewer (number of interviews the interviewer has performed, < 50, $50 - 100$, $100 - 200$, > 200), iv.mon (whether the interview was carried out during a month of harvest or planting, yes *vs.* no), iv.doa (whether the respondent has known someone who has died of AIDS yes *vs.* no). We assume $a = 3$, that is, at least 3 out of the five candidates are valid.

For all estimates, the data are weighted by survey weights. For individuals in group (a), HIV weights were used, for individuals in group (b), the individual survey weights were used; and for those in group (c), household survey weights were used. We examine HIV prevalence between genders and across different demographic, socio-economic and behavioural groups (Table S.4). Overall prevalence for women (16.1%) is higher than men (12.3%). In addition, this difference is consistent across all strata groups we examined. There are also significant differences among groups within a strata. For example, women at the lowest wealth quintile has a prevalence of only 8.8% compared to those in the highest two quintiles with over 20% prevalence.

Our results (Table S.5) show that across all scenarios, the imputation method gives very similar results to the unadjusted results. The partial identification bounds always give confidence intervals that are much wider. Between the two partial identification methods, the worst case is always less precise than the method proposed in this paper. The improved precision of the proposed method comes from a big reduction of the upper confidence interval. The improvement ranges from about 10% to 30%.

Table S.4: Observed proportions of HIV positive among the tested in 2007 Zambia DHS

		Women		Men	
		HIV prevalence	Number	HIV prevalence	Number
All		0.161	5713	0.123	5163
Age	15-19	0.058	1256	0.035	1109
	20-24	0.119	1119	0.053	830
	25-29	0.198	1102	0.115	772
	30-34	0.258	841	0.174	746
	35-39	0.25	588	0.223	594
	40-44	0.182	434	0.24	390
	45-49	0.122	373	0.183	318
Religion	Catholic	0.142	1086	0.116	1080
	Protestant	0.167	4525	0.125	3921
Location	Large city	0.247	450	0.205	439
	Small city	0.261	424	0.159	339
	Town	0.202	1662	0.126	1410
	Countryside	0.11	3177	0.095	2975
Wealth quintile	1st	0.088	903	0.07	938
	2nd	0.096	987	0.098	771
	3rd	0.133	1122	0.104	1044
	4th	0.229	1387	0.18	1271
	5th	0.216	1314	0.138	1139
Education	≤ 6	0.123	2516	0.085	1647
	> 6	0.191	3197	0.14	3516
Married	No	0.183	2269	0.083	2260
	Yes	0.147	3444	0.155	2901
Partners last 12m	0	0.153	1440	0.066	1223
	1	0.161	4188	0.127	3113
	1+	0.316	82	0.199	815
High risk sex last 12m	No	0.151	4903	0.118	3683
	Yes	0.231	807	0.136	1468
Condom use last sex	No	0.15	5141	0.111	4260
	Yes	0.268	569	0.179	898
STD last 12m	No	0.153	5428	0.113	4878
	Yes	0.339	264	0.306	269
Age first sex	Never	0.037	726	0.035	674
	≤ 15	0.173	1905	0.13	1652
	> 15	0.183	3080	0.141	2834
Sex last 12m	No	0.153	1440	0.066	1223
	Yes	0.164	4270	0.141	3928
Ever tested for HIV	No	0.132	3339	0.108	3945
	Yes	0.206	2349	0.173	1216

Table S.5: 95% confidence intervals for HIV prevalence estimates in 2007 Zambia DHS

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
All		LCI	0.152	0.159	0.117	0.12	0.123	0.115	0.186	0.083	0.089	0.086
		UCI	0.169	0.179	0.362	0.343	0.337	0.131	0.268	0.375	0.329	0.323
		Width	0.017	0.02	0.245	0.223	0.214	0.016	0.082	0.292	0.24	0.237
Age	15-19	LCI	0.038	0.058	0.029	0.029	0.024	0.019	0.049	0.014	0.015	0.013
		UCI	0.078	0.079	0.305	0.281	0.276	0.051	0.117	0.321	0.273	0.285
		Width	0.04	0.021	0.276	0.252	0.252	0.032	0.068	0.307	0.258	0.272
	20-24	LCI	0.099	0.117	0.073	0.075	0.077	0.036	0.108	0.026	0.027	0.025
		UCI	0.139	0.144	0.357	0.335	0.328	0.07	0.185	0.352	0.303	0.313
		Width	0.04	0.027	0.284	0.26	0.251	0.034	0.077	0.326	0.276	0.288
	25-29	LCI	0.179	0.189	0.14	0.143	0.145	0.095	0.187	0.065	0.071	0.083
		UCI	0.218	0.226	0.396	0.382	0.389	0.135	0.269	0.404	0.349	0.356
		Width	0.039	0.037	0.256	0.239	0.244	0.04	0.082	0.339	0.278	0.273
	30-34	LCI	0.237	0.237	0.186	0.19	0.202	0.153	0.284	0.11	0.118	0.123
		UCI	0.279	0.277	0.438	0.427	0.398	0.194	0.362	0.426	0.383	0.394
		Width	0.042	0.04	0.252	0.237	0.196	0.041	0.078	0.316	0.265	0.271
	35-39	LCI	0.226	0.213	0.176	0.181	0.181	0.198	0.317	0.144	0.155	0.155
		UCI	0.275	0.261	0.442	0.425	0.406	0.247	0.404	0.459	0.416	0.386
		Width	0.049	0.048	0.266	0.244	0.225	0.049	0.087	0.315	0.261	0.231
	40-44	LCI	0.155	0.188	0.122	0.124	0.135	0.213	0.318	0.164	0.172	0.172
		UCI	0.208	0.243	0.388	0.373	0.336	0.266	0.402	0.454	0.424	0.414
		Width	0.053	0.055	0.266	0.249	0.201	0.053	0.084	0.29	0.252	0.242
	45-49	LCI	0.095	0.122	0.075	0.077	0.077	0.152	0.26	0.114	0.122	0.129
		UCI	0.15	0.183	0.335	0.323	0.317	0.212	0.355	0.428	0.389	0.386
		Width	0.055	0.061	0.26	0.246	0.24	0.06	0.095	0.314	0.267	0.257

Table S.5: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Religion	Catholic	LCI	0.121	0.14	0.097	0.097	0.106	0.098	0.166	0.08	0.08	0.083
		UCI	0.164	0.177	0.34	0.332	0.312	0.134	0.258	0.311	0.303	0.289
		Width	0.043	0.037	0.243	0.235	0.206	0.036	0.092	0.231	0.223	0.206
	Protestant	LCI	0.159	0.16	0.127	0.128	0.137	0.116	0.198	0.092	0.093	0.098
		UCI	0.175	0.179	0.342	0.336	0.327	0.134	0.265	0.313	0.306	0.3
		Width	0.016	0.019	0.215	0.208	0.19	0.018	0.067	0.221	0.213	0.202
Residence	Large city	LCI	0.22	0.199	0.168	0.173	0.176	0.147	0.219	0.1	0.109	0.106
		UCI	0.243	0.229	0.425	0.406	0.4	0.171	0.303	0.443	0.39	0.372
		Width	0.023	0.03	0.257	0.233	0.224	0.024	0.084	0.343	0.281	0.266
	Small city	LCI	0.099	0.124	0.077	0.078	0.081	0.086	0.163	0.065	0.069	0.073
		UCI	0.121	0.141	0.32	0.304	0.304	0.105	0.231	0.325	0.287	0.293
		Width	0.022	0.017	0.243	0.226	0.223	0.019	0.068	0.26	0.218	0.22
Wealth quintile	1st	LCI	0.066	0.111	0.051	0.052	0.056	0.054	0.166	0.043	0.044	0.04
		UCI	0.111	0.136	0.316	0.294	0.306	0.086	0.234	0.284	0.257	0.258
		Width	0.045	0.025	0.265	0.242	0.25	0.032	0.068	0.241	0.213	0.218
	2nd	LCI	0.077	0.116	0.059	0.06	0.056	0.079	0.16	0.059	0.063	0.075
		UCI	0.115	0.136	0.328	0.314	0.3	0.118	0.236	0.361	0.315	0.338
		Width	0.038	0.02	0.269	0.254	0.244	0.039	0.076	0.302	0.252	0.263
	3rd	LCI	0.111	0.126	0.086	0.088	0.086	0.087	0.171	0.065	0.068	0.065
		UCI	0.155	0.146	0.353	0.336	0.32	0.121	0.238	0.361	0.321	0.321
		Width	0.044	0.02	0.267	0.248	0.234	0.034	0.067	0.296	0.253	0.256
	4th	LCI	0.211	0.181	0.163	0.168	0.174	0.165	0.213	0.115	0.127	0.114
		UCI	0.248	0.21	0.422	0.407	0.385	0.194	0.28	0.441	0.389	0.381
		Width	0.037	0.029	0.259	0.239	0.211	0.029	0.067	0.326	0.262	0.267
	5th	LCI	0.198	0.207	0.152	0.156	0.16	0.122	0.217	0.083	0.09	0.095
		UCI	0.235	0.244	0.417	0.398	0.395	0.154	0.309	0.432	0.38	0.355
		Width	0.037	0.037	0.265	0.242	0.235	0.032	0.092	0.349	0.29	0.26

Table S.5: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Education	≤ 6	LCI	0.112	0.114	0.084	0.087	0.085	0.071	0.142	0.052	0.054	0.055
		UCI	0.135	0.139	0.35	0.333	0.319	0.098	0.206	0.345	0.31	0.306
		Width	0.023	0.025	0.266	0.246	0.234	0.027	0.064	0.293	0.256	0.251
	> 6	LCI	0.179	0.186	0.14	0.145	0.146	0.13	0.206	0.094	0.101	0.096
		UCI	0.203	0.217	0.379	0.361	0.352	0.15	0.286	0.393	0.343	0.336
		Width	0.024	0.031	0.239	0.216	0.206	0.02	0.08	0.299	0.242	0.24
Married	No	LCI	0.17	0.176	0.136	0.137	0.141	0.07	0.13	0.056	0.056	0.059
		UCI	0.197	0.203	0.358	0.351	0.344	0.096	0.209	0.293	0.282	0.278
		Width	0.027	0.027	0.222	0.214	0.203	0.026	0.079	0.237	0.226	0.219
	Yes	LCI	0.136	0.142	0.108	0.109	0.118	0.144	0.241	0.116	0.117	0.121
		UCI	0.158	0.167	0.331	0.324	0.302	0.166	0.312	0.338	0.333	0.322
		Width	0.022	0.025	0.223	0.215	0.184	0.022	0.071	0.222	0.216	0.201
Partners last 12m	0	LCI	0.137	0.136	0.109	0.11	0.114	0.048	0.105	0.038	0.038	0.037
		UCI	0.169	0.173	0.348	0.34	0.34	0.085	0.187	0.303	0.294	0.282
		Width	0.032	0.037	0.239	0.23	0.226	0.037	0.082	0.265	0.256	0.245
	1	LCI	0.15	0.159	0.121	0.122	0.128	0.117	0.197	0.094	0.095	0.099
		UCI	0.171	0.183	0.338	0.332	0.311	0.136	0.277	0.319	0.313	0.299
		Width	0.021	0.024	0.217	0.21	0.183	0.019	0.08	0.225	0.218	0.2
	1+	LCI	0.253	0.206	0.231	0.229	0.264	0.176	0.254	0.152	0.153	0.162
		UCI	0.379	0.398	0.489	0.488	0.537	0.222	0.332	0.349	0.342	0.333
		Width	0.126	0.192	0.258	0.259	0.273	0.046	0.078	0.197	0.189	0.171
High risk sex last 12m	No	LCI	0.142	0.146	0.114	0.115	0.119	0.108	0.194	0.085	0.086	0.087
		UCI	0.159	0.166	0.331	0.325	0.314	0.127	0.26	0.32	0.315	0.309
		Width	0.017	0.02	0.217	0.21	0.195	0.019	0.066	0.235	0.229	0.222
	Yes	LCI	0.206	0.216	0.168	0.17	0.168	0.12	0.188	0.101	0.102	0.104
		UCI	0.255	0.277	0.404	0.395	0.379	0.152	0.26	0.304	0.293	0.282
		Width	0.049	0.061	0.236	0.225	0.211	0.032	0.072	0.203	0.191	0.178

Table S.5: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Condom use last sex	No	LCI	0.141	0.144	0.113	0.114	0.117	0.102	0.184	0.081	0.082	0.088
		UCI	0.16	0.165	0.333	0.326	0.314	0.121	0.25	0.307	0.3	0.291
		Width	0.019	0.021	0.22	0.212	0.197	0.019	0.066	0.226	0.218	0.203
	Yes	LCI	0.242	0.256	0.2	0.201	0.201	0.162	0.245	0.133	0.134	0.138
		UCI	0.294	0.329	0.426	0.42	0.393	0.197	0.33	0.36	0.353	0.336
		Width	0.052	0.073	0.226	0.219	0.192	0.035	0.085	0.227	0.219	0.198
STD last 12m	No	LCI	0.145	0.15	0.116	0.117	0.122	0.105	0.183	0.084	0.084	0.09
		UCI	0.161	0.172	0.333	0.327	0.314	0.121	0.248	0.308	0.301	0.291
		Width	0.016	0.022	0.217	0.21	0.192	0.016	0.065	0.224	0.217	0.201
	Yes	LCI	0.3	0.29	0.257	0.26	0.283	0.274	0.322	0.236	0.237	0.241
		UCI	0.378	0.404	0.476	0.47	0.424	0.339	0.445	0.452	0.442	0.448
		Width	0.078	0.114	0.219	0.21	0.141	0.065	0.123	0.216	0.205	0.207
Age first sex	Never	LCI	0.013	0.064	0.011	0.012	0.011	0.015	0.064	0.012	0.012	0.009
		UCI	0.06	0.099	0.273	0.267	0.268	0.055	0.157	0.308	0.295	0.297
		Width	0.047	0.035	0.262	0.255	0.257	0.04	0.093	0.296	0.283	0.288
	≤ 15	LCI	0.159	0.149	0.131	0.132	0.137	0.117	0.187	0.097	0.097	0.096
		UCI	0.187	0.169	0.339	0.334	0.32	0.144	0.268	0.303	0.297	0.298
		Width	0.028	0.02	0.208	0.202	0.183	0.027	0.081	0.206	0.2	0.202
	> 15	LCI	0.172	0.182	0.136	0.138	0.144	0.131	0.217	0.105	0.105	0.113
		UCI	0.194	0.208	0.363	0.356	0.342	0.152	0.298	0.332	0.327	0.312
		Width	0.022	0.026	0.227	0.218	0.198	0.021	0.081	0.227	0.222	0.199
Sex last 12m	No	LCI	0.137	0.139	0.109	0.11	0.114	0.048	0.107	0.038	0.038	0.037
		UCI	0.169	0.177	0.348	0.34	0.34	0.085	0.187	0.303	0.294	0.282
		Width	0.032	0.038	0.239	0.23	0.226	0.037	0.08	0.265	0.256	0.245
	Yes	LCI	0.155	0.162	0.124	0.125	0.132	0.132	0.217	0.108	0.108	0.114
		UCI	0.173	0.182	0.339	0.332	0.312	0.15	0.286	0.321	0.314	0.3
		Width	0.018	0.02	0.215	0.207	0.18	0.018	0.069	0.213	0.206	0.186
Ever tested for HIV	No	LCI	0.12	0.126	0.096	0.097	0.099	0.098	0.166	0.079	0.079	0.082
		UCI	0.144	0.147	0.323	0.318	0.301	0.118	0.236	0.305	0.298	0.288
		Width	0.024	0.021	0.227	0.221	0.202	0.02	0.07	0.226	0.219	0.206
	Yes	LCI	0.192	0.201	0.155	0.157	0.167	0.157	0.259	0.128	0.128	0.135
		UCI	0.22	0.234	0.374	0.367	0.359	0.19	0.341	0.353	0.347	0.348
		Width	0.028	0.033	0.219	0.21	0.192	0.033	0.082	0.225	0.219	0.213

S.6 HIV prevalence using 2004 Malawi DHS

The 2004 Malawi DHS (MDHS) was a national population-based survey carried out in two stages. In the first stage, 522 enumeration areas spread over 28 country districts were selected. Oversampling was done in some districts to provide more accurate district-level estimates. In the second stage, households were selected using systematic sampling. All women aged 15-49 in a selected household were eligible for interview.

In all the interviewed households, 12229 women (1733 urban and 10496 rural) were identified, and complete interviews were conducted with 11,698 women (95.7%). Every one in three households was selected for HIV testing. Within these households, 3797 men and 4071 women were eligible for HIV testing. Among the men, 2429 consented to testing, 759 refused, and the remaining 609 were un-contactable (mainly due to the moving). Testing was successfully conducted on 2404 men, resulting in a testing response rate of 63%. Failure to obtain test results was due to one of the following reasons: refusal, not being at home for the testing and unusable test results. For the purpose of the Jiang and Ding method, we used 3797 - 2404 - 759 = 634 as missing at random. Among the women, 2896 consented to testing, 916 refused (including those from parents). Testing was successfully conducted on 2864 women. For the purpose of the Jiang and Ding method, we used 4071 - 2864 - 916 = 291 as missing at random.

For the method proposed in this paper, we use the same list of instrumental variables: iv.lan (whether the language used in the questionnaire or interview is the same as the respondent's language, yes *vs.* no), iv.firstday (whether the interview was conducted on the first day of the interviews, yes *vs.* no), iv.interviewer (number of interviews the interviewer has performed, < 20, $20 - 50$, > 50 for women and < 50, > 50 for men), iv.mon (whether the interview was carried out during a month of harvest or planting, yes *vs.* no, in Malawi, the main harvesting season is April-July and planting is in October), iv.doa (whether the respondent knows someone who has died of AIDS, yes *vs.* no).

Our results (Table S.8) show that across all scenarios, the imputation method gives very similar results to the unadjusted results. The partial identification bounds always give confidence intervals that are much wider. Between the two partial identification methods, the worst case is always less precise than the method proposed in this paper. The improved precision of the proposed method comes from a big reduction of the upper confidence interval. The improvement ranges from about 10% to 30%.

Table S.6: Observed proportions of HIV positive among the tested in 2004 Malawi DHS

		Women		Men	
		HIV prevalence	Number	HIV prevalence	Number
All		0.133	2864	0.102	2404
Age	15-19	0.037	550	0.004	457
	20-24	0.13	691	0.034	424
	25-29	0.163	535	0.099	463
	30-34	0.163	392	0.204	354
	35-39	0.186	267	0.16	226
	40-44	0.189	244	0.199	225
	45-49	0.121	185	0.101	125
	50+	–	–	0.095	130
Religion	Catholic	0.138	633	0.102	531
	Presbyterian	0.097	495	0.089	440
	Anglican	0.177	55	0.052	45
	Adventist/baptist	0.121	174	0.166	149
	Other christian	0.138	1047	0.102	883
	Muslim	0.17	437	0.115	305
	None/Other	0.029	21	0.021	50
Residence	Urban	0.18	373	0.164	352
	Rural	0.125	2491	0.089	2052
Wealth quintile	1st	0.109	499	0.043	300
	2nd	0.103	578	0.047	494
	3rd	0.127	645	0.119	581
	4th	0.146	631	0.12	573
	5th	0.18	511	0.15	456
Education	≤ 6	0.125	1942	0.075	1283
	> 6	0.15	922	0.133	1121
Married	No	0.155	767	0.031	815
	Yes	0.125	2097	0.139	1589
Partners last 12m	0	0.124	578	0.047	462
	1	0.133	2265	0.111	1695
	1+	0.437	21	0.148	244
High risk sex last 12m	No	0.127	2676	0.105	1929
	Yes	0.218	188	0.087	472
Condom use last sex	No	0.129	2747	0.104	2129
	Yes	0.225	116	0.087	272
STD last 12m	No	0.121	2629	0.097	2272
	Yes	0.256	217	0.21	119
Age first sex	Never	0.025	255	0.018	242
	≤ 15	0.165	1027	0.088	649
	> 15	0.132	1578	0.122	1511
Sex last 12m	No	0.124	578	0.047	462
	Yes	0.136	2286	0.115	1939
Ever tested for HIV	No	0.129	2453	0.098	2004
	Yes	0.16	403	0.121	400

Table S.7: Relationship between HIV testing and some possible instrument variables

(a): Women

D		0		1	
Variable	N	Percent	N	Percent	Test
iv.lan	998		2855		X2= 7.419***
... 0	401	40.2%	1291	45.2%	
... 1	597	59.8%	1564	54.8%	
iv.firstday	1207		2864		X2= 0
... 0	747	61.9%	1775	62%	
... 1	460	38.1%	1089	38%	
iv.interviewer †	1001		2864		X2= 12.79***
... 0	81	8.1%	190	6.6%	
... 1	538	53.7%	1399	48.8%	
... 2	382	38.2%	1275	44.5%	
iv.mon	1001		2864		X2= 0.611
... 0	533	53.2%	1482	51.7%	
... 1	468	46.8%	1382	48.3%	
iv.doa	1001		2863		X2= 19.956***
... 0	423	42.3%	982	34.3%	
... 1	578	57.7%	1881	65.7%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

†0: ≤ 20; 1: $20 - 50$; 2: > 50

(b): Men

D		0		1	
Variable	N	Percent	N	Percent	Test
iv.lan	849		2398		X2= 9.146***
... 0	300	35.3%	991	41.3%	
... 1	549	64.7%	1407	58.7%	
iv.firstday	1393		2404		X2= 0.089
... 0	875	62.8%	1497	62.3%	
... 1	518	37.2%	907	37.7%	
iv.interviewer†	856		2404		X2= 0.288
... 0	63	7.4%	162	6.7%	
... 1	793	92.6%	2242	93.3%	
iv.mon	857		2404		X2= 3.453*
... 0	458	53.4%	1194	49.7%	
... 1	399	46.6%	1210	50.3%	
iv.doa	857		2404		X2= 38.756***
... 0	338	39.4%	671	27.9%	
... 1	519	60.6%	1733	72.1%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01
†0: ≤ 50; 1: > 50

Table S.8: 95% confidence intervals for HIV prevalence estimates in 2004 Malawi DHS

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
All		LCI	0.122	0.166	0.086	0.093	0.091	0.091	0.092	0.058	0.068	0.068
		UCI	0.144	0.295	0.402	0.355	0.365	0.113	0.113	0.444	0.345	0.345
		Width	0.022	0.129	0.316	0.262	0.274	0.022	0.021	0.386	0.277	0.277
Age	15-19	LCI	0.015	0.061	0.009	0.01	0.009	-0.019	0.006	-0.012	-0.014	-0.014
		UCI	0.06	0.162	0.398	0.327	0.328	0.027	0.021	0.433	0.328	0.314
		Width	0.045	0.101	0.389	0.317	0.319	0.046	0.015	0.445	0.342	0.328
	20-24	LCI	0.113	0.162	0.078	0.082	0.083	0.007	0.036	0.004	0.004	0.009
		UCI	0.147	0.306	0.413	0.378	0.376	0.061	0.062	0.432	0.318	0.288
		Width	0.034	0.144	0.335	0.296	0.293	0.054	0.026	0.428	0.314	0.279
	25-29	LCI	0.141	0.22	0.1	0.109	0.111	0.071	0.084	0.046	0.054	0.058
		UCI	0.185	0.358	0.43	0.384	0.387	0.128	0.12	0.445	0.353	0.354
		Width	0.044	0.138	0.33	0.275	0.276	0.057	0.036	0.399	0.299	0.296
	30-34	LCI	0.132	0.218	0.098	0.105	0.103	0.175	0.148	0.113	0.132	0.149
		UCI	0.195	0.338	0.423	0.382	0.388	0.234	0.199	0.526	0.434	0.422
		Width	0.063	0.12	0.325	0.277	0.285	0.059	0.051	0.413	0.302	0.273
	35-39	LCI	0.15	0.244	0.101	0.109	0.112	0.126	0.152	0.088	0.102	0.097
		UCI	0.222	0.353	0.486	0.446	0.435	0.195	0.216	0.48	0.384	0.399
		Width	0.072	0.109	0.385	0.337	0.323	0.069	0.064	0.392	0.282	0.302
	40-44	LCI	0.149	0.192	0.119	0.128	0.121	0.162	0.146	0.113	0.13	0.13
		UCI	0.229	0.317	0.418	0.378	0.368	0.236	0.211	0.501	0.42	0.426
		Width	0.08	0.125	0.299	0.25	0.247	0.074	0.065	0.388	0.29	0.296
	45-49	LCI	0.076	0.158	0.059	0.063	0.07	0.05	0.124	0.033	0.037	0.035
		UCI	0.166	0.302	0.387	0.348	0.347	0.152	0.176	0.501	0.421	0.426
		Width	0.09	0.144	0.328	0.285	0.277	0.102	0.052	0.468	0.384	0.391

Table S.8: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Religion	Catholic	LCI	0.112	0.177	0.088	0.09	0.092	0.075	0.078	0.062	0.06	0.078
		UCI	0.164	0.286	0.355	0.341	0.351	0.128	0.131	0.333	0.32	0.315
		Width	0.052	0.109	0.267	0.251	0.259	0.053	0.053	0.271	0.26	0.237
	Presbyterian	LCI	0.074	0.141	0.054	0.054	0.052	0.062	0.068	0.049	0.047	0.046
		UCI	0.12	0.27	0.368	0.36	0.361	0.117	0.116	0.356	0.355	0.32
		Width	0.046	0.129	0.314	0.306	0.309	0.055	0.048	0.307	0.308	0.274
	Anglican	LCI	0.085	0.203	0.057	0.054	0.047	-0.073	0.025	-0.027	-0.049	-0.033
		UCI	0.268	0.445	0.521	0.517	0.509	0.177	0.168	0.469	0.489	0.472
		Width	0.183	0.242	0.464	0.463	0.462	0.25	0.143	0.496	0.538	0.505
	Adventist/baptist	LCI	0.074	0.126	0.06	0.06	0.057	0.127	0.098	0.109	0.106	0.118
		UCI	0.168	0.297	0.377	0.369	0.343	0.204	0.196	0.391	0.392	0.343
		Width	0.094	0.171	0.317	0.309	0.286	0.077	0.098	0.282	0.286	0.225
	Other christian	LCI	0.12	0.179	0.089	0.09	0.093	0.082	0.088	0.063	0.061	0.068
		UCI	0.156	0.296	0.381	0.368	0.379	0.122	0.125	0.358	0.356	0.351
		Width	0.036	0.117	0.292	0.278	0.286	0.04	0.037	0.295	0.295	0.283
	Muslim	LCI	0.142	0.208	0.102	0.105	0.117	0.081	0.076	0.055	0.052	0.065
		UCI	0.198	0.384	0.438	0.412	0.426	0.15	0.146	0.442	0.436	0.42
		Width	0.056	0.176	0.336	0.307	0.309	0.069	0.07	0.387	0.384	0.355
	None/Other	LCI	-0.138	0.044	-0.1	-0.108	-0.089	-0.057	0.005	-0.029	-0.043	-0.041
		UCI	0.195	0.474	0.576	0.56	0.596	0.099	0.132	0.509	0.525	0.504
		Width	0.333	0.43	0.676	0.668	0.685	0.156	0.127	0.538	0.568	0.545
Residence	Large city	LCI	0.152	0.196	0.098	0.107	0.102	0.137	0.102	0.074	0.095	0.103
		UCI	0.208	0.297	0.496	0.448	0.452	0.191	0.127	0.565	0.447	0.443
		Width	0.056	0.101	0.398	0.341	0.35	0.054	0.025	0.491	0.352	0.34
	Small city	LCI	0.112	0.181	0.08	0.086	0.085	0.077	0.087	0.05	0.058	0.059
		UCI	0.137	0.288	0.391	0.344	0.358	0.1	0.113	0.422	0.331	0.334
		Width	0.025	0.107	0.311	0.258	0.273	0.023	0.026	0.372	0.273	0.275

Table S.8: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Wealth quintile	1st	LCI	0.083	0.188	0.057	0.061	0.06	0.015	0.069	0.008	0.01	0.015
		UCI	0.134	0.291	0.408	0.362	0.368	0.072	0.089	0.469	0.36	0.356
		Width	0.051	0.103	0.351	0.301	0.308	0.057	0.02	0.461	0.35	0.341
	2nd	LCI	0.078	0.159	0.054	0.058	0.056	0.023	0.077	0.016	0.017	0.024
		UCI	0.127	0.298	0.388	0.342	0.362	0.071	0.1	0.414	0.319	0.317
		Width	0.049	0.139	0.334	0.284	0.306	0.048	0.023	0.398	0.302	0.293
	3rd	LCI	0.102	0.158	0.073	0.08	0.089	0.099	0.093	0.067	0.077	0.081
		UCI	0.152	0.285	0.394	0.346	0.361	0.139	0.12	0.432	0.34	0.315
		Width	0.05	0.127	0.321	0.266	0.272	0.04	0.027	0.365	0.263	0.234
	4th	LCI	0.124	0.173	0.089	0.096	0.102	0.098	0.096	0.067	0.077	0.079
		UCI	0.168	0.273	0.409	0.367	0.365	0.143	0.119	0.436	0.34	0.351
		Width	0.044	0.1	0.32	0.271	0.263	0.045	0.023	0.369	0.263	0.272
	5th	LCI	0.158	0.202	0.104	0.113	0.112	0.124	0.097	0.068	0.082	0.087
		UCI	0.203	0.302	0.481	0.436	0.44	0.176	0.129	0.547	0.449	0.437
		Width	0.045	0.1	0.377	0.323	0.328	0.052	0.032	0.479	0.367	0.35
Education	≤ 6	LCI	0.113	0.164	0.079	0.085	0.082	0.06	0.07	0.036	0.042	0.045
		UCI	0.137	0.281	0.405	0.358	0.366	0.091	0.095	0.448	0.349	0.352
		Width	0.024	0.117	0.326	0.273	0.284	0.031	0.025	0.412	0.307	0.307
	> 6	LCI	0.133	0.205	0.095	0.103	0.1	0.116	0.11	0.076	0.089	0.089
		UCI	0.167	0.319	0.411	0.363	0.376	0.15	0.139	0.452	0.353	0.346
		Width	0.034	0.114	0.316	0.26	0.276	0.034	0.029	0.376	0.264	0.257
Married	No	LCI	0.136	0.196	0.099	0.101	0.103	0.01	0.02	0.009	0.007	0.016
		UCI	0.175	0.329	0.412	0.395	0.412	0.052	0.044	0.336	0.331	0.312
		Width	0.039	0.133	0.313	0.294	0.309	0.042	0.024	0.327	0.324	0.296
	Yes	LCI	0.114	0.164	0.086	0.087	0.085	0.126	0.127	0.097	0.096	0.097
		UCI	0.137	0.268	0.359	0.346	0.353	0.152	0.157	0.369	0.364	0.362
		Width	0.023	0.104	0.273	0.259	0.268	0.026	0.03	0.272	0.268	0.265

Table S.8: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Partners last 12m	0	LCI	0.102	0.164	0.074	0.074	0.075	0.024	0.027	0.017	0.016	0.019
		UCI	0.145	0.282	0.389	0.376	0.371	0.069	0.06	0.369	0.363	0.336
		Width	0.043	0.118	0.315	0.302	0.296	0.045	0.033	0.352	0.347	0.317
	1	LCI	0.122	0.189	0.091	0.094	0.091	0.1	0.1	0.076	0.075	0.08
		UCI	0.144	0.284	0.368	0.353	0.358	0.122	0.13	0.357	0.35	0.342
		Width	0.022	0.095	0.277	0.259	0.267	0.022	0.03	0.281	0.275	0.262
	1+	LCI	0.231	0.294	0.152	0.138	0.217	0.109	0.109	0.092	0.09	0.099
		UCI	0.642	0.677	0.773	0.78	0.733	0.187	0.189	0.386	0.375	0.372
		Width	0.411	0.383	0.621	0.642	0.516	0.078	0.08	0.294	0.285	0.273
High risk sex last 12m	No	LCI	0.117	0.17	0.086	0.088	0.087	0.093	0.096	0.07	0.069	0.069
		UCI	0.138	0.295	0.367	0.352	0.361	0.118	0.12	0.356	0.351	0.347
		Width	0.021	0.125	0.281	0.264	0.274	0.025	0.024	0.286	0.282	0.278
	Yes	LCI	0.175	0.251	0.133	0.135	0.137	0.061	0.062	0.047	0.046	0.055
		UCI	0.26	0.39	0.459	0.447	0.447	0.113	0.106	0.36	0.355	0.347
		Width	0.085	0.139	0.326	0.312	0.31	0.052	0.044	0.313	0.309	0.292
Condom use last sex	No	LCI	0.119	0.179	0.088	0.09	0.088	0.092	0.089	0.069	0.069	0.072
		UCI	0.14	0.308	0.369	0.354	0.368	0.117	0.113	0.354	0.35	0.347
		Width	0.021	0.129	0.281	0.264	0.28	0.025	0.024	0.285	0.281	0.275
	Yes	LCI	0.17	0.227	0.148	0.147	0.162	0.047	0.072	0.038	0.035	0.034
		UCI	0.279	0.381	0.429	0.425	0.415	0.128	0.141	0.377	0.372	0.38
		Width	0.109	0.154	0.281	0.278	0.253	0.081	0.069	0.339	0.337	0.346
STD last 12m	No	LCI	0.109	0.166	0.082	0.082	0.082	0.086	0.086	0.064	0.064	0.068
		UCI	0.133	0.278	0.362	0.349	0.358	0.108	0.11	0.348	0.345	0.343
		Width	0.024	0.112	0.28	0.267	0.276	0.022	0.024	0.284	0.281	0.275
	Yes	LCI	0.217	0.26	0.17	0.176	0.19	0.156	0.133	0.132	0.125	0.16
		UCI	0.296	0.441	0.479	0.455	0.453	0.263	0.257	0.447	0.447	0.435
		Width	0.079	0.181	0.309	0.279	0.263	0.107	0.124	0.315	0.322	0.275

Table S.8: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Age first sex	Never	LCI	-0.014	0.064	-0.01	-0.012	-0.013	-0.022	0.007	-0.013	-0.017	-0.021
		UCI	0.064	0.182	0.373	0.363	0.359	0.059	0.027	0.395	0.391	0.369
		Width	0.078	0.118	0.383	0.375	0.372	0.081	0.02	0.408	0.408	0.39
	≤ 15	LCI	0.146	0.186	0.114	0.116	0.116	0.065	0.085	0.053	0.05	0.068
		UCI	0.183	0.296	0.374	0.363	0.363	0.11	0.108	0.335	0.33	0.327
		Width	0.037	0.11	0.26	0.247	0.247	0.045	0.023	0.282	0.28	0.259
	> 15	LCI	0.117	0.196	0.086	0.088	0.088	0.109	0.106	0.083	0.081	0.086
		UCI	0.147	0.326	0.382	0.367	0.375	0.134	0.134	0.363	0.36	0.354
		Width	0.03	0.13	0.296	0.279	0.287	0.025	0.028	0.28	0.279	0.268
Sex last 12m	No	LCI	0.102	0.152	0.074	0.074	0.075	0.024	0.028	0.017	0.016	0.019
		UCI	0.145	0.293	0.389	0.376	0.371	0.069	0.062	0.369	0.363	0.336
		Width	0.043	0.141	0.315	0.302	0.296	0.045	0.034	0.352	0.347	0.317
	Yes	LCI	0.124	0.194	0.093	0.094	0.093	0.104	0.105	0.078	0.078	0.083
		UCI	0.147	0.304	0.367	0.353	0.358	0.127	0.131	0.352	0.348	0.342
		Width	0.023	0.11	0.274	0.259	0.265	0.023	0.026	0.274	0.27	0.259
Ever tested for HIV	No	LCI	0.117	0.172	0.087	0.089	0.087	0.086	0.084	0.064	0.063	0.068
		UCI	0.14	0.293	0.369	0.354	0.363	0.11	0.112	0.356	0.353	0.347
		Width	0.023	0.121	0.282	0.265	0.276	0.024	0.028	0.292	0.29	0.279
	Yes	LCI	0.133	0.205	0.098	0.1	0.102	0.094	0.106	0.076	0.073	0.082
		UCI	0.187	0.329	0.412	0.397	0.404	0.149	0.171	0.356	0.355	0.35
		Width	0.054	0.124	0.314	0.297	0.302	0.055	0.065	0.28	0.282	0.268

S.7 HIV prevalence using 2003 Kenya DHS

In the 2003 Kenya DHS (KDHS), all men were selected for individual questionnaires, and all women in the same household were approached for an HIV test. In total, there were 4183 eligible men and 4303 eligible women. HIV test was successfully conducted on 2941 of the eligible men. In the remaining eligible men, 545 refused testing, 512 were not contactable, and 185 were not for other reasons. For eligible women, these numbers are 3283, 620, 257, and 141, respectively. For the Jiang and Ding method, we used 512+ 185 = 697 and 257+141 = 398 as missing at random for the men and women.

The observed HIV prevalence was 8.7% among women (n = 3283) and 4.7% among men (n = 2941) (Table S.9). The age-specific analysis revealed variations in HIV prevalence, with higher rates observed among older age groups. For instance, the 50+ age group had an HIV prevalence of 5.8% among men. Additionally, urban areas showed higher HIV prevalence (12.3%) compared to rural areas (7.5%). The relationship between HIV testing and several potential instrumental variables is presented in Table refAtable7.2. We found significant associations between HIV testing and all instrument variables except iv.interviewer for women and iv.doa for men.

For the method proposed in this paper, we use the same list of instrumental variables: iv.lan (whether the language used in the questionnaire or interview is the same as the respondent's language, yes *vs.* no), iv.firstday (whether the interview was conducted on the first day of the interviews, yes *vs.* no), iv.interviewer (number of interviews the interviewer has performed, < 50, $50-100$, > 100 for women and < 50, > 50 for men), iv.mon (whether the interview was carried out during a month of harvest or planting, yes *vs.* no, in Kenya, 90% of the farmers grow maize, and the main harvesting season is March-April and October-December), iv.doa (whether the respondent knows someone who has died of AIDS, yes *vs.* no).

For both men and women, the width of the confidence intervals varies across different methods. The Marra method generally produces wider intervals compared to the Unadjusted method, indicating increased uncertainty when considering correlated data. The Worst case method often has the widest intervals, while JD and IV methods generally yield narrower intervals compared to the Worst case and Marra methods, reflecting their potential for more precise estimates. The widths vary considerably across demographic and behavioural subgroups.

Our results (Table S.11) show that the imputation method gives very similar results to the unadjusted results across all scenarios. The partial identification bounds always give confidence intervals that are much wider. Between the two partial identification methods, the worst case is always less precise than the method proposed in this paper. The improved precision of the proposed method comes from a big reduction in the upper confidence interval. The improvement ranges from about 10% to 30%.

Table S.9: Observed proportions of HIV positive among the tested in 2004 Kenya DHS

		Women		Men	
		HIV prevalence	Number	HIV prevalence	Number
All		0.087	3283	0.047	2941
Age	15-19	0.032	735	0.004	697
	20-24	0.089	681	0.019	550
	25-29	0.126	545	0.088	416
	30-34	0.115	478	0.054	358
	35-39	0.118	350	0.098	298
	40-44	0.097	296	0.087	257
	45-49	0.043	198	0.047	173
	50+	–	0	0.058	192
Religion	Catholic	0.089	776	0.047	749
	Protestant	0.092	2074	0.047	1686
	Muslim	0.027	357	0.029	301
	None/Other	0.111	54	0.053	174
Residence	Urban	0.123	985	0.078	856
	Rural	0.075	2298	0.037	2085
Wealth quintile	1st	0.038	561	0.039	479
	2nd	0.086	588	0.04	502
	3rd	0.072	602	0.025	523
	4th	0.097	640	0.045	614
	5th	0.122	892	0.074	823
Education	≤ 6	0.074	1260	0.039	962
	> 6	0.094	2023	0.051	1979
Married	No	0.098	1287	0.024	1410
	Yes	0.08	1984	0.069	1507
Partners last 12m	0	0.058	953	0.014	801
	1	0.096	2259	0.053	1794
	1+	0.21	55	0.086	321
High risk sex last 12m	No	0.075	2875	0.045	2153
	Yes	0.172	392	0.05	763
Condom use last sex	No	0.084	3141	0.047	2571
	Yes	0.153	124	0.04	344
STD last 12m	No	0.084	3140	0.044	2819
	Yes	0.19	118	0.146	77
Age first sex	Never	0.016	539	0.009	457
	≤ 15	0.122	956	0.051	1085
	> 15	0.088	1771	0.054	1365
Sex last 12m	No	0.058	953	0.014	801
	Yes	0.099	2314	0.058	2115
Ever tested for HIV	No	0.081	2785	0.042	2456
	Yes	0.122	485	0.071	461

Table S.10: Relationship between HIV testing and some possible instrument variables

(a): Women

D		0		1	
Variable	N	Percent	N	Percent	Test
iv.lan	772		3270		X2= 3.614*
... 0	394	51%	1542	47.2%	
... 1	378	49%	1728	52.8%	
iv.firstday	1090		3333		X2= 7.45***
... 0	600	55%	1674	50.2%	
... 1	490	45%	1659	49.8%	
iv.interviewer	772		3271		X2= 0.693
... 0	136	17.6%	575	17.6%	
... 1	577	74.7%	2416	73.9%	
... 2	59	7.6%	280	8.6%	
iv.mon	772		3271		X2= 31.212***
... 0	610	79%	2845	87%	
... 1	162	21%	426	13%	
iv.doa	767		3270		X2= 8.051***
... 0	237	30.9%	843	25.8%	
... 1	530	69.1%	2427	74.2%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

†0: ≤ 50; 1: $50 - 100$; 2: > 100

(b): Men

D		0		1	
Variable	N	Percent	N	Percent	Test
iv.lan	661		2915		X2= 12.931***
... 0	340	51.4%	1272	43.6%	
... 1	321	48.6%	1643	56.4%	
iv.firstday	1352		3025		X2= 12.093***
... 0	744	55%	1491	49.3%	
... 1	608	45%	1534	50.7%	
iv.interviewer	661		2917		X2= 0.041
... 0	25	3.8%	118	4%	
... 1	636	96.2%	2799	96%	
iv.mon	661		2917		X2= 28.131***
... 0	525	79.4%	2551	87.5%	
... 1	136	20.6%	366	12.5%	
iv.doa	659		2917		X2= 0.578
... 0	177	26.9%	739	25.3%	
... 1	482	73.1%	2178	74.7%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01
†0: ≤ 50; 1: > 50

Table S.11: 95% confidence intervals for HIV prevalence estimates in 2004 Kenya DHS

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
All		LCI	0.081	0.104	0.061	0.068	0.067	0.042	0.042	0.029	0.036	0.033
		UCI	0.098	0.181	0.323	0.243	0.267	0.054	0.087	0.352	0.207	0.229
		Width	0.017	0.077	0.262	0.175	0.2	0.012	0.045	0.323	0.171	0.196
Age	15-19	LCI	0.02	0.033	0.016	0.017	0.02	-0.01	0.006	-0.007	-0.008	-0.008
		UCI	0.05	0.1	0.299	0.205	0.239	0.017	0.026	0.265	0.138	0.171
		Width	0.03	0.067	0.283	0.188	0.219	0.027	0.02	0.272	0.146	0.179
	20-24	LCI	0.073	0.099	0.057	0.063	0.061	0.007	0.016	0.005	0.005	0.008
		UCI	0.104	0.194	0.322	0.243	0.274	0.03	0.057	0.37	0.222	0.242
		Width	0.031	0.095	0.265	0.18	0.213	0.023	0.041	0.365	0.217	0.234
	25-29	LCI	0.112	0.128	0.085	0.096	0.094	0.072	0.051	0.046	0.059	0.058
		UCI	0.149	0.23	0.368	0.277	0.306	0.102	0.14	0.444	0.271	0.281
		Width	0.037	0.102	0.283	0.181	0.212	0.03	0.089	0.398	0.212	0.223
	30-34	LCI	0.094	0.13	0.072	0.079	0.083	0.042	0.06	0.029	0.036	0.044
		UCI	0.141	0.246	0.367	0.283	0.314	0.072	0.147	0.388	0.222	0.242
		Width	0.047	0.116	0.295	0.204	0.231	0.03	0.087	0.359	0.186	0.198
	35-39	LCI	0.094	0.121	0.074	0.08	0.085	0.084	0.062	0.056	0.069	0.062
		UCI	0.143	0.251	0.358	0.287	0.307	0.121	0.153	0.441	0.291	0.341
		Width	0.049	0.13	0.284	0.207	0.222	0.037	0.091	0.385	0.222	0.279
	40-44	LCI	0.076	0.097	0.058	0.061	0.067	0.066	0.058	0.047	0.056	0.056
		UCI	0.126	0.198	0.377	0.307	0.321	0.105	0.172	0.395	0.262	0.268
		Width	0.05	0.101	0.319	0.246	0.254	0.039	0.114	0.348	0.206	0.212
	45-49	LCI	0.029	0.053	0.022	0.022	0.024	0.023	0.033	0.015	0.017	0.018
		UCI	0.069	0.175	0.34	0.281	0.287	0.077	0.106	0.415	0.274	0.29
		Width	0.04	0.122	0.318	0.259	0.263	0.054	0.073	0.4	0.257	0.272

Table S.11: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Religion	Catholic	LCI	0.072	0.092	0.061	0.063	0.062	0.035	0.034	0.03	0.03	0.036
		UCI	0.106	0.202	0.277	0.231	0.27	0.06	0.093	0.243	0.19	0.217
		Width	0.034	0.11	0.216	0.168	0.208	0.025	0.059	0.213	0.16	0.181
	Protestent	LCI	0.081	0.096	0.067	0.07	0.064	0.038	0.04	0.032	0.034	0.036
		UCI	0.102	0.18	0.274	0.227	0.266	0.057	0.093	0.235	0.177	0.226
		Width	0.021	0.084	0.207	0.157	0.202	0.019	0.053	0.203	0.143	0.19
	Muslim	LCI	0.014	0.058	0.013	0.01	0.014	0.013	0.024	0.011	0.008	0.011
		UCI	0.063	0.299	0.314	0.288	0.29	0.068	0.173	0.359	0.305	0.34
		Width	0.049	0.241	0.301	0.278	0.276	0.055	0.149	0.348	0.297	0.329
	None/Other	LCI	0.078	0.049	0.052	0.052	0.057	0.025	0.022	0.023	0.02	0.025
		UCI	0.144	0.239	0.41	0.401	0.4	0.081	0.123	0.294	0.255	0.285
		Width	0.066	0.19	0.358	0.349	0.343	0.056	0.101	0.271	0.235	0.26
Residence	Large city	LCI	0.113	0.11	0.072	0.086	0.082	0.066	0.039	0.037	0.051	0.047
		UCI	0.138	0.253	0.449	0.34	0.389	0.087	0.156	0.481	0.289	0.334
		Width	0.025	0.143	0.377	0.254	0.307	0.021	0.117	0.444	0.238	0.287
	Small city	LCI	0.067	0.092	0.055	0.059	0.058	0.031	0.035	0.024	0.026	0.027
		UCI	0.088	0.16	0.265	0.206	0.221	0.046	0.072	0.29	0.175	0.178
		Width	0.021	0.068	0.21	0.147	0.163	0.015	0.037	0.266	0.149	0.151
Wealth quintile	1st	LCI	0.023	0.084	0.025	0.024	0.022	0.026	0.041	0.022	0.023	0.028
		UCI	0.065	0.143	0.199	0.149	0.168	0.056	0.084	0.28	0.186	0.18
		Width	0.042	0.059	0.174	0.125	0.146	0.03	0.043	0.258	0.163	0.152
	2nd	LCI	0.067	0.093	0.061	0.061	0.065	0.027	0.037	0.022	0.025	0.035
		UCI	0.104	0.147	0.241	0.199	0.2	0.055	0.075	0.267	0.152	0.171
		Width	0.037	0.054	0.18	0.138	0.135	0.028	0.038	0.245	0.127	0.136
	3rd	LCI	0.056	0.076	0.047	0.051	0.054	0.014	0.027	0.011	0.012	0.014
		UCI	0.091	0.158	0.279	0.209	0.235	0.038	0.065	0.309	0.184	0.189
		Width	0.035	0.082	0.232	0.158	0.181	0.024	0.038	0.298	0.172	0.175
	4th	LCI	0.08	0.092	0.062	0.069	0.072	0.032	0.028	0.024	0.028	0.036
		UCI	0.112	0.177	0.333	0.26	0.283	0.061	0.066	0.34	0.214	0.244
		Width	0.032	0.085	0.271	0.191	0.211	0.029	0.038	0.316	0.186	0.208
	5th	LCI	0.112	0.12	0.068	0.083	0.078	0.062	0.047	0.035	0.047	0.047
		UCI	0.137	0.257	0.467	0.358	0.396	0.082	0.131	0.478	0.287	0.328
		Width	0.025	0.137	0.399	0.275	0.318	0.02	0.084	0.443	0.24	0.281

Table S.11: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Education	≤ 6	LCI	0.065	0.101	0.051	0.056	0.068	0.029	0.037	0.021	0.025	0.033
		UCI	0.092	0.175	0.302	0.234	0.248	0.052	0.088	0.324	0.19	0.203
		Width	0.027	0.074	0.251	0.178	0.18	0.023	0.051	0.303	0.165	0.17
	> 6	LCI	0.086	0.101	0.064	0.072	0.07	0.044	0.036	0.03	0.037	0.04
		UCI	0.105	0.19	0.342	0.253	0.288	0.059	0.086	0.368	0.219	0.252
		Width	0.019	0.089	0.278	0.181	0.218	0.015	0.05	0.338	0.182	0.212
Married	No	LCI	0.086	0.102	0.068	0.073	0.073	0.013	0.021	0.012	0.012	0.014
		UCI	0.111	0.192	0.301	0.247	0.295	0.034	0.057	0.22	0.156	0.209
		Width	0.025	0.09	0.233	0.174	0.222	0.021	0.036	0.208	0.144	0.195
	Yes	LCI	0.072	0.078	0.059	0.062	0.061	0.06	0.053	0.05	0.051	0.052
		UCI	0.092	0.168	0.26	0.226	0.25	0.081	0.122	0.266	0.22	0.247
		Width	0.02	0.09	0.201	0.164	0.189	0.021	0.069	0.216	0.169	0.195
Partners last 12m	0	LCI	0.044	0.058	0.035	0.037	0.038	0	0.012	0.002	0.001	0
		UCI	0.072	0.14	0.286	0.226	0.286	0.03	0.056	0.234	0.164	0.223
		Width	0.028	0.082	0.251	0.189	0.248	0.03	0.044	0.232	0.163	0.223
	1	LCI	0.09	0.113	0.075	0.078	0.074	0.045	0.044	0.038	0.039	0.045
		UCI	0.106	0.208	0.27	0.231	0.256	0.064	0.096	0.241	0.192	0.226
		Width	0.016	0.095	0.195	0.153	0.182	0.019	0.052	0.203	0.153	0.181
	1+	LCI	0.172	0.148	0.149	0.154	0.148	0.067	0.051	0.056	0.056	0.054
		UCI	0.248	0.376	0.369	0.345	0.428	0.105	0.169	0.3	0.249	0.272
		Width	0.076	0.228	0.22	0.191	0.28	0.038	0.118	0.244	0.193	0.218
High risk sex last 12m	No	LCI	0.068	0.09	0.055	0.058	0.056	0.038	0.036	0.032	0.033	0.033
		UCI	0.085	0.157	0.264	0.221	0.256	0.054	0.095	0.244	0.19	0.241
		Width	0.017	0.067	0.209	0.163	0.2	0.016	0.059	0.212	0.157	0.208
	Yes	LCI	0.146	0.167	0.125	0.131	0.13	0.038	0.036	0.033	0.034	0.036
		UCI	0.198	0.308	0.355	0.307	0.355	0.062	0.09	0.232	0.178	0.206
		Width	0.052	0.141	0.23	0.176	0.225	0.024	0.054	0.199	0.144	0.17

Table S.11: Continued

			Women					Men				
			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
Condom use last sex	No	LCI	0.077	0.1	0.064	0.067	0.066	0.041	0.038	0.034	0.036	0.035
		UCI	0.094	0.172	0.269	0.227	0.265	0.055	0.09	0.236	0.183	0.232
		Width	0.017	0.072	0.205	0.16	0.199	0.014	0.052	0.202	0.147	0.197
	Yes	LCI	0.121	0.134	0.101	0.106	0.099	0.022	0.03	0.02	0.019	0.022
		UCI	0.185	0.302	0.372	0.305	0.354	0.057	0.092	0.263	0.21	0.244
		Width	0.064	0.168	0.271	0.199	0.255	0.035	0.062	0.243	0.191	0.222
STD last 12m	No	LCI	0.077	0.1	0.063	0.066	0.065	0.037	0.035	0.031	0.033	0.031
		UCI	0.093	0.181	0.273	0.228	0.265	0.052	0.086	0.239	0.183	0.231
		Width	0.016	0.081	0.21	0.162	0.2	0.015	0.051	0.208	0.15	0.2
	Yes	LCI	0.154	0.166	0.142	0.147	0.139	0.1	0.112	0.098	0.091	0.098
		UCI	0.226	0.318	0.331	0.293	0.343	0.193	0.307	0.31	0.298	0.319
		Width	0.072	0.152	0.189	0.146	0.204	0.093	0.195	0.212	0.207	0.221
Age first sex	Never	LCI	-0.005	0.038	-0.003	-0.004	-0.002	-0.006	0.006	-0.003	-0.006	-0.004
		UCI	0.036	0.105	0.264	0.196	0.262	0.026	0.036	0.238	0.165	0.226
		Width	0.041	0.067	0.267	0.2	0.264	0.032	0.03	0.241	0.171	0.23
	≤ 15	LCI	0.112	0.13	0.095	0.098	0.098	0.039	0.038	0.034	0.035	0.037
		UCI	0.137	0.208	0.283	0.251	0.272	0.063	0.09	0.228	0.181	0.214
		Width	0.025	0.078	0.188	0.153	0.174	0.024	0.052	0.194	0.146	0.177
	> 15	LCI	0.077	0.105	0.063	0.066	0.064	0.045	0.046	0.037	0.039	0.042
		UCI	0.102	0.194	0.283	0.239	0.274	0.066	0.115	0.266	0.208	0.266
		Width	0.025	0.089	0.22	0.173	0.21	0.021	0.069	0.229	0.169	0.224
Sex last 12m	No	LCI	0.044	0.063	0.035	0.037	0.038	0	0.015	0.002	0.001	0
		UCI	0.072	0.142	0.286	0.226	0.286	0.03	0.065	0.234	0.164	0.223
		Width	0.028	0.079	0.251	0.189	0.248	0.03	0.05	0.232	0.163	0.223
	Yes	LCI	0.091	0.117	0.075	0.079	0.076	0.052	0.044	0.043	0.045	0.044
		UCI	0.111	0.199	0.272	0.235	0.259	0.067	0.108	0.245	0.197	0.229
		Width	0.02	0.082	0.197	0.156	0.183	0.015	0.064	0.202	0.152	0.185
Ever tested for HIV	No	LCI	0.075	0.092	0.062	0.064	0.064	0.035	0.035	0.029	0.03	0.033
		UCI	0.091	0.18	0.263	0.222	0.258	0.051	0.083	0.23	0.179	0.218
		Width	0.016	0.088	0.201	0.158	0.194	0.016	0.048	0.201	0.149	0.185
	Yes	LCI	0.103	0.124	0.081	0.087	0.08	0.056	0.045	0.045	0.049	0.052
		UCI	0.141	0.259	0.342	0.281	0.314	0.087	0.139	0.299	0.221	0.289
		Width	0.038	0.135	0.261	0.194	0.234	0.031	0.094	0.254	0.172	0.237