5-2024

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

Received: 20 April 2023 | Revised: 21 April 2024 | Accepted: 29 April 2024

DOI: 10.1002/sim.10108

RESEARCH ARTICLE

Statistics in Medicine WILEY

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, Casuarina, Australia

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

Correspondence
Oyelola A. Adegboye, Menzies School of Health Research, Charles Darwin University, Casuarina, 0810, Darwin, Australia.
Email: oyelola.adegboye@menzies.edu.au

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.

KEYWORDS

demographic and health surveys, HIV, instrumental variable, non-response, partial identification

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 Marmarange 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 logistic regression to predict the HIV prevalence among the non-respondents under a non-informative non-response assumption in a 12-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 is comparable. Reniers et al,[8] Bärnighausen et al,[10] Hogan et al[11] adjusted non-response bias by a Heckman-type selection model,[12] 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[13] constructed bounds based on the partial identification approach of Menski.[14,15] 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[13] used restrictions implied by the dynamics of HIV (i.e, 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.[16] 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 identify the quantity of interest correctly. Following Manski,[17] Chernozhukov et al,[18] Windmeijer et al[19] 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[14] 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 Reference 17, $\forall v \in \mathcal{V}$,

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

The inequalities imply

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

$$\leq \inf_{v\in\mathcal{V}} \{\mu_{1v}\mathrm{P}(D = 1|V = v) + K_1\mathrm{P}(D = 0|V = v)\} \equiv \mathrm{UB}_V, \tag{6}$$

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, for example, Lawlor et al[16] and Kreider.[20] 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:

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

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 E(Y). For any two sets of bounds that both include $E(Y)$, their intersection must be non-empty, and also correctly identify 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 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 (LB, UB), can be empirically estimated as $(\widehat{\text{LB}}, \widehat{\text{UB}})$. Constructing a confidence interval involves considering an approximate $b_0 \times 100$ percent interval for $(\widehat{\text{LB}} - z_{(1-b_0)/2}\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{UB}} + z_{(1-b_0)/2}\widehat{\text{SE}}_{\text{UB}})$, 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. However, to address the width issue for μ,[21] alternative bounds are suggested, $(\widehat{\text{LB}} - C_n\widehat{\text{SE}}_{\text{LB}}, \widehat{\text{UB}} + C_n\widehat{\text{SE}}_{\text{UB}})$, 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[22] 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 STUDIES

3.1 | Simulation study 1

We use a set of simulation studies 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 $b_1 = (b_{11}, \ldots, b_{1L})^T$ 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: $b_1 = (\overbrace{0, \ldots, 0}^{s}, \overbrace{1, \ldots, 1}^{(L-s)})^T$; and $b_1 = (\overbrace{0, \ldots, 0}^{s}, \overbrace{4, \ldots, 4}^{(L-s)})^T$. 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 $c_1 = (c_{11}, \ldots, c_{1L})^T$ give the association between each instrument and D. We consider two situations, (a) Strong instruments: $c_1 = (5, \ldots, 5)$ 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 $c_{11}, \ldots, c_{1L}$. 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 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 1,000 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.[23] 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, that is, 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 1,000 simulations, the approximate 95% confidence intervals include E(Y). The second row gives the lower confidence limits, averaged over 1,000 simulations. The third row gives the upper confidence limits, averaged over 1,000 simulations. The fourth row gives the average width of the confidence intervals.

TABLE 1 Partial identification of E(Y) with $L = 5$ instruments and $s = 3$ valid instruments with E(Y) fixed at 0.15.

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

Note: Results are stratified by average non-response rate $1 - E(D) = 0.1$ 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.

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 in 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[24] and Jiang and Ding.[25] The simulations follow the approach of Clark and Houle[26] 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 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 $n = 7{,}000$. 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,[24] 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 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,[24] we use only interviewer as the IV. We try a selection of representative copulas: Normal, Frank, Clayton rotated 90°, and Clayton rotated 270° 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[24]).

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

We consider eight scenarios defined by the combination of the interviewer's IV strength (weak/strong) and the number of invalid IVs (cases (1)–(4)). For each scenario, we use 1,000 simulations to compare the following methods: (1) standard imputation; (2) the selection method of Marra et al[24] 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[25] 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 six methods. Table 4 shows the widths of the 95% confidence intervals using the six 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 a Normal copula (the correct one under the setup of the simulation study) or the best, adjusts the observed prevalence upwards. When interviewer IV is invalid (case (1)), the Marra confidence intervals give overestimates 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, the selection is based on the effectiveness of each interviewer (see

TABLE 2 Summary statistics of eight 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 eight scenarios considered in simulation study 2.

Interviewer	# Invalid	Imputation		Marra Normal Copula		Marra Best		Worst case bounds		JD bounds		Union bounds	
Effects	IVs	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 eight 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

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 ≥ JD ≥ 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 7,146 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, 7,406 females were eligible for interview and HIV testing, while the number of eligible males was 7,146. 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, 1,695 (22.9%) of the women and 1,983 (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 of those who are not tested to adjust for potential biases.[27] 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, informed by previous studies.[28,29] 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

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	5,713	0.123	5,163
15–19	0.058	1,256	0.035	1,109
20–24	0.119	1,119	0.053	830
25–29	0.198	1,102	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.

D	0		1		
Variable	**N**	**Percent**	**N**	**Percent**	**Test**
(a) Women					
iv.lan	1,695		5,713		$\chi^2 = 24.706^{***}$
… 0	727	42.9%	2,845	49.8%	
… 1	968	57.1%	2,868	50.2%	
iv.firstday	1,695		5,713		$\chi^2 = 22.774^{***}$
… 0	1,094	64.5%	3,315	58%	
… 1	601	35.5%	2,398	42%	
iv.interviewer	1,433		5,713		$\chi^2 = 28.929^{***}$
… 0	74	5.2%	256	4.5%	
… 1	190	13.3%	807	14.1%	
… 2	695	48.5%	2,369	41.5%	
… 3	474	33.1%	2,281	39.9%	
iv.mon	1,695		5,713		$\chi^2 = 82.634^{***}$
… 0	311	18.3%	1,688	29.5%	
… 1	1,384	81.7%	4,025	70.5%	
iv.doa	1,695		5,713		$\chi^2 = 14.097^{***}$
… 0	688	40.6%	2,616	45.8%	
… 1	1,007	59.4%	3,097	54.2%	
(b) Men					
iv.lan	1,983		5,163		$\chi^2 = 160.186^{***}$
… 0	739	37.3%	2,789	54%	
… 1	1,244	62.7%	2,374	46%	
iv.firstday	1,983		5,163		$\chi^2 = 4.267^{**}$
… 0	1,251	63.1%	3,118	60.4%	
… 1	732	36.9%	2,045	39.6%	
iv.interviewer	1339		5,161		$\chi^2 = 13.453^{***}$
… 0	167	12.5%	583	11.3%	
… 1	62	4.6%	373	7.2%	
… 2	493	36.8%	1,791	34.7%	
… 3	617	46.1%	2,414	46.8%	
iv.mon	1,983		5,163		$\chi^2 = 200.984^{***}$
… 0	293	14.8%	1,621	31.4%	
… 1	1,690	85.2%	3,542	68.6%	
iv.doa	1,983		5,163		$\chi^2 = 60.931^{***}$
… 0	650	32.8%	2,216	42.9%	
… 1	1,333	67.2%	2,947	57.1%	

Note: Statistical significance markers: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

TABLE 7 The 95% confidence intervals for HIV prevalence estimates in 2007 Zambia DHS.

			Unadjusted	Marra	Worst case	JD	IV	Unadjusted	Marra	Worst case	JD	IV
			(a) Women					(b) Men				
All		LCI	0.153	0.159	0.118	0.121	0.123	0.116	0.186	0.084	0.089	0.086
		UCI	0.169	0.179	0.361	0.343	0.339	0.13	0.268	0.374	0.328	0.322
		Width	0.016	0.02	0.243	0.222	0.216	0.014	0.082	0.29	0.239	0.236
Age	15–19	LCI	0.038	0.058	0.029	0.029	0.023	0.019	0.049	0.014	0.015	0.013
		UCI	0.078	0.189	0.305	0.282	0.274	0.05	0.187	0.319	0.271	0.279
		Width	0.04	0.131	0.276	0.253	0.251	0.031	0.138	0.305	0.256	0.266
	20–24	LCI	0.099	0.058	0.073	0.076	0.076	0.035	0.049	0.025	0.026	0.024
		UCI	0.139	0.189	0.36	0.336	0.331	0.072	0.187	0.351	0.302	0.315
		Width	0.04	0.131	0.287	0.26	0.255	0.037	0.138	0.326	0.276	0.291
	25–29	LCI	0.181	0.058	0.142	0.144	0.147	0.096	0.049	0.067	0.073	0.083
		UCI	0.216	0.189	0.392	0.378	0.383	0.134	0.187	0.409	0.353	0.361
		Width	0.035	0.131	0.25	0.234	0.236	0.038	0.138	0.342	0.28	0.278
	30–34	LCI	0.237	0.058	0.187	0.191	0.202	0.155	0.049	0.111	0.12	0.124
		UCI	0.279	0.189	0.44	0.429	0.401	0.192	0.187	0.426	0.382	0.384
		Width	0.042	0.131	0.253	0.238	0.199	0.037	0.138	0.315	0.262	0.26
	35–39	LCI	0.223	0.058	0.174	0.179	0.178	0.201	0.049	0.147	0.158	0.16
		UCI	0.278	0.189	0.443	0.425	0.408	0.244	0.187	0.46	0.417	0.376
		Width	0.055	0.131	0.269	0.246	0.23	0.043	0.138	0.313	0.259	0.216
	40–44	LCI	0.151	0.058	0.12	0.121	0.131	0.215	0.049	0.166	0.173	0.173
		UCI	0.212	0.189	0.391	0.377	0.342	0.264	0.187	0.452	0.421	0.414
		Width	0.061	0.131	0.271	0.256	0.211	0.049	0.138	0.286	0.248	0.241
	45–49	LCI	0.09	0.058	0.072	0.073	0.073	0.153	0.049	0.115	0.122	0.129
		UCI	0.155	0.189	0.333	0.322	0.315	0.212	0.187	0.424	0.386	0.377
		Width	0.065	0.131	0.261	0.249	0.242	0.059	0.138	0.309	0.264	0.248

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.

The worst case method presents the widest CIs, indicating a conservative approach. Methods such as Marra, JD, and IV provide moderate CI widths, pointing to a balance between raw data and adjusted estimates. IV typically shows a slightly tighter range than JD. Demographic variables influence the variability in HIV prevalence estimates, with age and socioeconomic status being particularly pronounced. For men aged 45–49, the worst case CI width is 0.314, substantially wider than the unadjusted width of 0.06, highlighting increased uncertainty in older populations. The socioeconomic dimension is reflected in the wealth quintile analysis, where the highest quintiles exhibit larger CI widths in the worst case method, such as 0.349 for the top quintile.

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 Zambia DHS. 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%.

For both men and women, the width of the confidence intervals varies across different methods. The worst case method consistently shows the widest CIs. The Marra, JD, and IV methods produce moderate CI widths, with the IV method slightly more conservative than JD. 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 the JD and IV methods generally yield narrower intervals compared to the worst case and Marra methods, reflecting their potential for more concise estimates. The widths vary considerably across demographic and behavioural subgroups.

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.

The 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, for example, interviewing process, interviewer characteristics, and so forth, 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,[24] Manski's[21] worst case bounds, and Jiang and Ding.[25] 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.[11,30] 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 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.[31,32] Our method is similar to that proposed in Kang et al[33] for estimating causal effects when some instruments are possibly invalid. Kang et al[33] 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,[33] 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 our method is to create identification bounds prevalence that is 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 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.[34] 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.* Fortunately, the strength of an instrument is a testable hypothesis. We demonstrated how this could be done in the empirical study.

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. Open access publishing facilitated by Charles Darwin University, as part of the Wiley - Charles Darwin University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declared no conflict of interest for this article.

DATA AVAILABILITY STATEMENT

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

ENDNOTE

*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 one is valid.

ORCID

Oyelola A. Adegboye https://orcid.org/0000-0002-9793-8024

REFERENCES

1. WHO. *Global HIV/AIDS Response: Epidemic Update and Health Sector Progress towards Universal Access: Progress Report 2011*. Geneva: WHO; 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 J, Gouws E, Ghys PD. National population-based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. *Sex Transm Infect*. 2006;82(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. *Sex Transm Infect*. 2008;84(1):i71-i77.
5. Mishra V, Barrere B, Hong R, Khan S. Evaluation of bias in HIV seroprevalence estimates from national household surveys. *Sex Transm Infect*. 2008;84(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. *Stat Methods Med Res*. 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. Tech. Rep*. Paris: Université Paris Descartes, INED, IRD; 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, 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. 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.
11. 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. *Sex Transm Infect*. 2012;88(2):i17-i23.
12. Heckman JJ. Sample selection bias as a specification error. *Econometrica*. 1979;47:153-161.
13. 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. *J R Stat Soc A*. 2014;177:587-606.
14. Manski CF. Anatomy of the selection problem. *J Human Resour*. 1989;24(3):343-360.
15. Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev*. 1990;80(2):319-323.
16. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey SG. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163. doi:10.1002/sim.3034
17. Manski CF. *Partial Identification of Probability Distributions*. New York: Springer-Verlag; 2003.
18. Chernozhukov V, Lee S, Rosen AM. Intersection bounds: estimation and inference. *Econometrica*. 2013;81(2):667-737. doi:10.3982/ECTA8718
19. Windmeijer F, Liang X, Hartwig FP, Bowden J. The confidence interval method for selecting valid instrumental variables. *J R Stat Soc Series B Stat Methodology*. 2021;83(4):752-776. doi:10.1111/rssb.12449

20. Kreider B, Hill SC. Partially identifying treatment effects with an application to covering the uninsured. *J Human Resour*. 2009;44:409-449.
21. Imbens G, Manski CF. Confidence intervals for partially identified parameters. *Econometrica*. 2004;72:1845-1857.
22. Horowitz JL, Manski CF. Identification and estimation of statistical functionals using incomplete data. *J Econometr*. 2006;132:445-459.
23. WHO/UNAIDS. *Guidelines for Measuring National HIV Prevalence in Population-Based Surveys*. Geneva: WHO; 2005.
24. Marra G, Radice R, Bärnighausen T, Wood SN, McGovern ME. A simultaneous equation approach to estimating HIV prevalence with nonignorable missing responses. *J Am Stat Assoc*. 2017;112(518):484-496.
25. Jiang Z, Ding P. Using missing types to improve partial identification with application to a study of HIV prevalence in Malawi. *Ann Appl Stat*. 2018;12(3):1831-1852.
26. Clark S, 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. Cent. Stat. Soc. Sci. Work. Pap. no 120, University of Washington. 2012.
27. 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. Tech. Rep*. Calverton, Maryland, USA: Macro International; 2008.
28. Adegboye OA, Ezechukwu HC, Woodall H, et al. Media exposure, behavioural risk factors and HIV testing among women of reproductive age in Papua New Guinea: a cross-sectional study. *Tropic Med Infect Disease*. 2022;7(2):30.
29. Kareem YO, Dorgbetor CI, Ameyaw EK, et al. Assessment and associated factors of comprehensive HIV knowledge in an at-risk population: a cross-sectional study from 19,286 young persons in Nigeria. *Therap Adv Infect Disease*. 2023;10:20499361231163664.
30. Nicoletti C. Poverty analysis with missing data: alternative estimators compared. *Empir Econ*. 2010;38:1-22.
31. 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 Transm Infect*. 2003;79(6):442-447. doi:10.1136/sti.79.6.442
32. Genberg BL, Hlavka Z, Konda KA, 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. *Soc Sci Med*. 2009;68(12):2279-2287. doi:10.1016/j.socscimed.2009.04.005
33. 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. doi:10.1111/biom.13415
34. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat*. 2002;20:518-529.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Adegboye OA, Fujii T, Leung DH-Y, Siyu L. HIV estimation using population-based surveys with non-response: A partial identification approach. *Statistics in Medicine*. 2024;1-15. doi: 10.1002/sim.10108