Measuring Uncertainty in a Standard Classification Procedure of Population-based HIV Assessment (PHIA) Survey Results

Jared Garfinkel

Supervisors: Dr. Stephen Delgado, Giles Reid, Dr. Ying Wei

Organization of Practicum: ICAP at Columbia, PHIA Project

I. Overview and Student Role

ICAP at Columbia is a global health organization headquartered at Columbia University Irving Medical Center. One of their public health goals is to achieve full treatment and suppression of HIV/AIDS for people living with HIV (PLWH) ("HIV/AIDS" 2020). The Joint United Nations Programme on HIV/AIDS (UNAIDS) tracks a target-based approach called the 90-90-90 goals that aims to measure the percent of PLWH who are 1) aware of their HIV+ status, 2) adhering to antiretroviral therapy (ARV), and 3) virally suppressed ("90-90-90"). The goals state that 90% of those who are HIV+ should be aware of their status; 90% of those who are aware of their status should be adhering to ARVs; and, 90% of those who are adhering to ARVs should be virally suppressed.

The Population-based HIV Impact Assessment (PHIA) Survey measures self-reported and blood-test criteria to determine the level of treatment. The PHIA Team at ICAP at Columbia uses this data to determine the status of individual respondents in regards to their awareness, ARV adherence, and viral load suppression. This data is then reported to track progress towards the 90-90-90 goals of 90 percent of PLWH in each target area.

From June 2020 I took part in a field work exercise at ICAP at Columbia, a global health organization with offices at Columbia University Irving Medical Center, as a graduate student assistant. I worked closely with my practicum mentor, Dr. Stephen Delgado, and my academic advisor, Dr. Ying Wei, to whom I am grateful for their guidance and mentorship. I was also fortunate to have the assistance of Giles Reid, our team statistician.

II. Background

PHIA surveys are in-depth surveys of the living situations of PLWH and typically includes a blood test except in circumstances where a respondent does not agree to consent to draw blood

for some reason. The results of the PHIA survey are used for reporting purposes to track the progress of the UNAIDS 90-90-90 goals in participating countries. It is of interest to researchers to find the sensitivity and specificity of the reported results to gain insight into the data collected by the survey as well as testing behaviors of respondents.

How can PHIA survey results be confirmed? In the absence of a gold standard, a classification procedure was adopted to determine the percent of PLWH who were aware of their HIV+ status, on ARV, and virally suppressed. The process used self-reported response items and blood tests, when available, to create a final determination of each respondent's status in each category. In other words, a respondent who self-reports that they are unaware of their HIV+ status will be reclassified as "aware" for reporting purposes if that individual tests positive for ARVs. It follows that if a person is taking or has taken ARVs recently that they have accepted treatment for HIV or AIDS and therefore must have known their disease status.

Furthermore, based on this classification procedure, individuals who self-reported that they are aware of their HIV positive status are not reclassified as unaware regardless of the presence of ARVs in their blood. The reclassification is only in one direction – from unaware status to aware status based on the blood test results for ARVs. This will always result in no change or an increase in the number of respondents classified as aware of their HIV positive status, but never results in a reduction of those who are aware of their HIV positive status.

Blood tests can be used to measure ARV status and viral load. There is an interest among researchers to determine if viral load can be used as a proxy for ARV status. Antiretroviral detection blood tests are typically costly in terms of time and expense. By comparison the viral load blood tests are faster and cheaper to use. Young, et. al (2020) conducted reclassification procedures based on self-report responses only compared to those with ARV adjustment, viral suppression (<550 copies/mL), and both. As would be expected, the treatment coverage increased with adjustments from ARV adjustment, viral suppression, and both. However, this study was limited by its sample size (n=305, 205, and 150 for awareness status, ARV adherence, and viral load suppression, respectively).

Latent Class Analysis (LCA) is a cluster-based analysis that can be used for diagnostic assessment with no gold standard. Using binary outcome variables, latent class analysis is an unsupervised learning technique used to classify observations by outcome. It can be used to

measure unobserved classes of interest. Unsupervised learning is a clustering algorithm that uses observed outcomes to categorize observations into distinct groups. It is desirable that latent classes are easily interpretable based on the observations that were grouped together. The usefulness of LCA depends on the interpretability of the classes.

LCA is a finite mixture model usually used with multivariate categorical data.

A mixture model uses the theory that any single continuously distributed variable can be approximated by a mixture of several normally distributed variables.

Let $X = (x_1, \ldots, x_p)$ be a response vector for individual i to p categorical (ordered or nominal) items.

Let c represent the underlying latent class variable with K categories.

Let w_k = Pr(c = k) represent the probability of being in the kth latent class in the sample such that $\sum_{k=1}^{K} w_k = 1$

Let π_{jk} = Pr($x_j | c = k$) $j = 1, \ldots, p$ $k = 1, \ldots, K$ be the probability of endorsing item j if a respondent is in class k.

The LCA model is a mixture model with the following parameters:

Pr(x) = $\sum_{k=1}^{K} w_k \prod_{j=1}^{p} \pi_{jk}$

Breakout Box 1. Details for latent class analysis.

Building an LCA Model

What gets included? The process to select the dataset for the latent class analysis started with choosing the observations that are eligible. The eligibility requirements were that a respondent had a positive HIV status and that they had given consent to have blood tests.

Then the outcome variables in the survey were removed if they were not related to the classes of interest, being awareness, ARV adherence, and viral load suppression. Ultimately, out of 926 items, 7 dummy variables were created from 14 outcome variables (Table 1). The expected classes that we considered were “unaware”, “aware – not on ART”, and “aware – on ART”. These classes are desirable because they would quantify the unobserved true underlying status of the “first 90” and “second 90” in the 90-90-90 goals.

Differences between countries may change the performance of the model. In countries with lower infection rates, there may be fewer PLWH in the study. Furthermore, where fewer respondents are likely to be aware of their HIV status, adhering to ARVs or virally suppressed there is likely to be higher variability in the data.

III. Methods

Read in Data

Population-based HIV Impact Assessment Survey data is collected by partners across 15 countries. Blood tests are used for ARV detection of anti-retroviral therapy and viral load detection. The results are sent to one or more central locations in each country to be collected by a storage administrator.

A process was used to create the dataset used for LCA. First, the data was limited to only those respondents who were reported to be living with HIV and above the age of consent of 15 years old. Then covariates of interest were selected based on their clinical relevance to the 90-90-90 status including ARV blood test results and viral load test results. These variables included biological and self-reported survey data.

LCA uses binary data to cluster observations by class. So, all the dummy variables and blood tests were coded into binary variables. LCA works by clustering outcome variables from measurable outcomes into unobservable latent classes, in this case, the true 90-90-90 status of each respondent.

Perform LCA

How many classes in the LCA? The number of classes is limited by the number of covariates included in the model. There are methods for determining the reliability and validity of the model including AIC or BIC. How can the classes be interpreted? The classes can be interpreted based on summaries of the observations in the features they create.

What characterizes those in each class? The item probabilities in each class can be used to interpret group membership. Each item has a probability of being endorsed in a group depending on how many of the observations in a group endorsed that item. These results can be compared to what we expected.

Is there a subgroup or response block of interest? We are particularly interested in examples of respondents who claim to have no knowledge of their HIV positive status while accepting treatment for ARVs. However, we are interested in measuring the sensitivity and specificity of those who are classified as aware and those who are on ARVs compared to the underlying truth. Who is likely to report being unaware or tested negative while having detectable ARVs? We also asked about the testing behaviors and characteristics of respondents who were non-disclosing of their aware status or ARV adherence. Who is likely to report being unaware while being viral

load suppressed (<200mg/mL)? Similarly, those who are virally suppressed are assumed to be on ARVs. One point of discussion is the likelihood of an “elite suppressor” who can be virally suppressed and not on ARVs, however, for the purposes of reporting, these instances are considered negligible.

Cross-Tabulations

This information can be presented as tables of outcome probabilities by extracting class data and then joining datasets for cross tabulation. This is useful to evaluate sensitivity and specificity. In this study, confusionMatrix in R was used to create summary classes using current classification as the truth. While the objective is prediction of the current classification, the underlying truth is also of clinical interest.

Cross tabulations were used to perform exploratory data analysis and to measure the sensitivity and specificity of the LCA output using the reported PHIA results as the truth. The results that PHIA reports are the summary statistics from each country using the final determination of the 90-90-90 status for all eligible respondents.

There were essentially 3 different datasets used to capture the unobserved classes. All 3 datasets included self-reported items that measured the past testing and self-reported positive or negative test results and whether a respondent took ARVs or received other care for HIV. In addition, 1 dataset used the results of an ARV detection blood test, and another dataset included a cutoff variable of viral load suppression based on the results of viral load blood test. In this way, we tested a “base” dataset compared with 2 datasets that each included 1 blood test. This analysis was conducted for 3 country-level datasets.

How do results compare between countries? As mentioned earlier, there can be wide variability depending on the prevalence of HIV in the study population and the percent of individuals who are aware, on ARVs, and virally suppressed.

Model building

The process for building the model started with the outcome variables in the PHIA survey. The measured observations were classified as their final outcome in each of the three categories –

awareness, ARV adherence, and viral load suppression. However, this model did not allow for the expected classes and were difficult to interpret.

There were challenges related to the complete rate of some variables. Dummy variables were used to account for responses to several questions that had low rates of diagnostically significant items.

Variable Name	Description	Source
reported_positive	The respondent has endorsed awareness during the interview.	Self report (Questionnaire)
never_tested	The respondent has not endorsed having been tested for HIV/AIDS.	Self report (Questionnaire)
reported_negative	The respondent has not endorsed awareness during the interview.	Self report (Questionnaire)
arvstakenev	The respondent has endorsed having taken ARVs.	Self report (Questionnaire)
oth_art_care	The respondent has endorsed receiving care for HIV/AIDS	Self report (Questionnaire
arvscurrent	The respondent has endorsed currently taking ARVs	Self report (Questionnaire)
vlunder200	The respondent has less than 200 copies of virus per milliliter of blood	Lab Results
arvstatus	The respondent tested positive for ARVs	Lab Results

Table 1. Outcome variables.

IV. Results

The final model contained 6 self-reported variables and 2 lab results. A 4-class latent class analysis was used to improve interpretability and to capture expected groups. Tables were used to show the characteristics of observations in each class.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care	vlunder200
1	0.0000	0.9999	1e-04	0.0000	0.0000	0.0000	0.0487
2	0.0000	0.0383	1e+00	1.0000	0.9928	0.9799	0.1541
3	0.0000	0.0703	1e+00	0.0051	0.0000	0.4817	0.0469
4	0.7882	0.0000	0e+00	0.0000	0.0000	0.0000	0.0202

Figure 1. In Lesotho, for the model using viral load cutoff of 200 copies per millileter, classes would apparently be: 1 – unaware, reported negative; 2 – aware, on art; 3 – aware, not on art; and, 4 -never tested.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care	arvstatus
1	0.0000	0.0386	1.0000	1.0000	0.9989	0.9798	0.9618
2	0.0000	0.9360	0.0019	0.0000	0.0000	0.0000	0.1647
3	0.0000	0.0655	1.0000	0.0494	0.0000	0.5049	0.0818
4	0.9999	0.0000	0.0000	0.0000	0.0000	0.0000	0.0769

Figure 2. In Lesotho, for the model using ARV detection results (present or absent), the resulting classes would be: 1 – aware, on art; 2 – unaware, reported negative; 3 – aware, not on art; and, 4 – never tested.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care
1	0.0000	0.0383	1e+00	1.000	0.9927	0.9798
2	0.8988	0.0000	0e+00	0.000	0.0000	0.0000
3	0.0000	0.0704	1e+00	0.004	0.0000	0.4814
4	0.0000	0.9618	1e-04	0.000	0.0000	0.0000

Figure 3. In Lesotho, the model without lab data specifies classes as follows: 1 – aware, on art; 2 – never tested; 3 – aware, not on art; and, 4 – unaware, reported negative.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction aware unaware
##    aware     2647       0
##    unaware    127     704
##
##                Accuracy : 0.9635
##                  95% CI : (0.9567, 0.9695)
##     No Information Rate : 0.7976
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.894
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9542
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.8472
##              Prevalence : 0.7976
##          Detection Rate : 0.7611
##    Detection Prevalence : 0.7611
##       Balanced Accuracy : 0.9771
##
##        'Positive' Class : aware
##
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   on art not on art
##   on art       2339          0
##   not on art    143        292
##
##                Accuracy : 0.9484
##                  95% CI : (0.9396, 0.9564)
##     No Information Rate : 0.8947
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.775
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9424
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.6713
##              Prevalence : 0.8947
##          Detection Rate : 0.8432
##    Detection Prevalence : 0.8432
##       Balanced Accuracy : 0.9712
##
##        'Positive' Class : on art
##
```

Figure 4. In Lesotho, the model using the ARV status as an outcome had a high sensitivity and specificity for awareness (sensitivity = 0.9542, specificity = 1) and for ART adherence (sensitivity = 0.9424, specificity = 1) using the PHIA-reported results as the truth.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care	vlunder200
1	0.0000	0.0271	0.9998	0	0.0000	0.1622	0.0271
2	0.0000	0.0440	1.0000	1	0.9633	1.0000	0.0856
3	0.7753	0.0000	0.0000	0	0.0000	0.0000	0.0121
4	0.0000	0.9998	0.0000	0	0.0000	0.0000	0.0316

Figure 5. In Cameroon, the classes for the model with the viral load cutoff variable can be described as follows: 1 – aware, not on art; 2 – aware, on art; 3 – never tested; 4 – unaware, reported negative.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care	arvstatus
1	0.9999	0.0000	0	0.0000	0	0.0000	0.1347
2	0.0000	0.0385	1	0.2885	0	0.4038	0.2116
3	0.0000	0.0434	1	1.0000	1	1.0000	0.9158
4	0.0000	0.8349	0	0.0000	0	0.0000	0.1829

Figure 6. In Cameroon, the classes for the model with ARV detection were as follows: 1 – never tested; 2 – aware, not on art; 3 – aware, on art; and, 4 – unaware, reported negative.

Class	never_tested	reported_negative	reported_positive	arvstakenev	arvscurrent	oth_art_care
1	0.0000	0.9019	0.0000	0	0.0000	0.0000
2	0.0000	0.0440	1.0000	1	0.9633	1.0000
3	0.0000	0.0271	0.9998	0	0.0000	0.1622
4	0.8854	0.0000	0.0000	0	0.0000	0.0000

Figure 7. In Cameroon, the model without any lab data returned the following classes: 1 – unaware, reported negative; 2 – aware, on art; 3 – aware, not on art; and, 4 – never tested.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction aware unaware
##    aware     439       0
##    unaware    88     443
##
##                Accuracy : 0.9093
##                  95% CI : (0.8894, 0.9266)
##     No Information Rate : 0.5433
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.82
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.8330
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.8343
##              Prevalence : 0.5433
##          Detection Rate : 0.4526
##    Detection Prevalence : 0.4526
##       Balanced Accuracy : 0.9165
##
##        'Positive' Class : aware
##
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   on art not on art
##   on art        392          0
##   not on art     99         36
##
##                Accuracy : 0.8121
##                  95% CI : (0.7761, 0.8446)
##     No Information Rate : 0.9317
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.3511
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.7984
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.2667
##              Prevalence : 0.9317
##          Detection Rate : 0.7438
##    Detection Prevalence : 0.7438
##       Balanced Accuracy : 0.8992
##
##        'Positive' Class : on art
##
```

Figure 8. In Cameroon, the model using ARV detection lab data had a sensitivity of 0.8330 and specificity of 1 for awareness, and for ART adherence, sensitivity was 0.7984 and specificity was 1 using the PHIA-reported results as the truth.

V. Conclusions/ Discussion

This project is part of an ongoing multi-country study with complex data security and human blood sample handling systems in 15 countries for sensitive personal health information. The scope was narrowed to the model building component for the LCA using survey data and biological data from previous studies of 3 countries. Ultimately, it may not be entirely useful to compare the latent classes to those that resulted from the current classification process, since the objective of LCA is to determine a true underlying status for each respondent. However, this was a useful exercise as it resulted in a process that can be modified in the future to meet the needs of the project.

Limitations

There were 3 assumptions made. 1) ARV detection is the de facto gold standard for measuring awareness of HIV-status. Those who were receiving treatment for HIV were considered to have been aware of their status at the time of their interview. 2) Those who self-reported that they are aware are also considered to be aware regardless of ARV detection. 3) Missing values and non-endorsement of items were treated as 0s for simplicity. This may bias the results since non-disclosure may occur when the truth is that respondent was in fact aware or on ARVs.

Is this problematic? The results are considered to be conservative since a response other than endorsement of an item is treated as absent. Misclassification based on ARV status is unidirectional, from unaware to aware. Differential misclassification results in bias towards the null in hypothesis testing. So, the odds of being reported as having a 90-90-90 status given the covariates is possibly lower than the true value.

Future Research

This research continues by implementing Bayesian latent class analysis with simulation methods in JAGS. Bayesian latent class analysis allows for model based simulations by updating prior values for parameters of interest. Simulation methods are useful when direct estimation is computationally intensive.

Further research will focus on sensitivity and specificity analysis and testing behavior in all the PHIA countries. The motivation of the project is to understand the potential effect of using viral load blood tests as a proxy for ARV detection blood tests.

VI. References

Xia, Yiqing, et al. "Accuracy of Self-Reported HIV Testing History and Awareness of HIV-Positive Status among People Living with HIV in Four Sub-Saharan African Countries." *AIDS*, 2020, doi:10.1101/2020.09.16.20196105.

Young, Peter W., et al. "Use of Viral Load to Improve Survey Estimates of Known HIV-Positive Status and Antiretroviral Treatment Coverage." *AIDS*, vol. 34, no. 4, 2020, pp. 631–636., doi:10.1097/qad.0000000000002453.

Diggle, Peter J. "Estimating Prevalence Using an Imperfect Test." *Epidemiology Research International*, vol. 2011, 2 Aug. 2011, pp. 1–5., doi:10.1155/2011/608719.

Rindskopf, David, and Wallace Rindskopf. "The Value of Latent Class Analysis in Medical Diagnosis." *Statistics in Medicine*, vol. 5, no. 1, 1986, pp. 21–27., doi:10.1002/sim.4780050105.

Qu, Yinsheng, et al. "Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostic Tests." *Biometrics*, vol. 52, no. 3, 1996, p. 797., doi:10.2307/2533043.

Patterson, Blossom H, et al. "Latent Class Analysis of Complex Sample Survey Data." *Journal of the American Statistical Association*, vol. 97, no. 459, 2002, pp. 721–741., doi:10.1198/016214502388618465.

Beath, Ken J. "RandomLCA: An R Package for Latent Class with Random Effects Analysis." *Journal of Statistical Software*, vol. 81, no. 13, 2017, doi:10.18637/jss.v081.i13.

"HIV/AIDS." *ICAP at Columbia University*, Columbia University Mailman School of Public Health, 15 Nov. 2020, icap.columbia.edu/our-focus-areas/hiv-aids/.

"90-90-90: Treatment for All." *UNAIDS*, UNAIDS, www.unaids.org/en/resources/909090.