

**Sampling Simulation Programs to Explore Outcome Variability in the Sampling Plans Proposed by UCLA and the Kinshasa School of Public Health**

**Part 1 – Non-LQAS Sampling Plans**

Dale Rhoda  
April 15, 2023

Table of Content

[Introduction 3](#_Toc132378155)

[Methods 3](#_Toc132378156)

[Defining how the outcome is distributed among the target population 3](#_Toc132378157)

[Simulation sampling plans 6](#_Toc132378158)

[Results 8](#_Toc132378159)

[Discussion 13](#_Toc132378160)

[Next Steps 14](#_Toc132378161)

# Introduction

A recent UCLA – Kinshasa School of Public Health (KSPH) draft protocol suggests that a vaccination coverage survey with five primary sampling units (PSUs or clusters) per stratum should be investigated as a generalizable alternative to the designs in the WHO 2018 Vaccination Coverage Cluster Survey Reference Manual, which recommends using at least 30 PSUs for precise estimation and at least 15 to classify coverage into broad outcome categories (1).

I have concerns about using a small number of clusters in strata where outcomes can be spatially heterogeneous – with immunization programs performing at a high level in some portions of a district and performing more poorly elsewhere within the same district (or stratum). My work supporting Bill & Melinda Gates Foundation (BMGF) surveys designed to be representative at the health district level in both Burkina Faso (2016) and Pakistan (2020-2022) indicates that when coverage is not near 0% or 100%, it tends to vary widely across clusters, even within a district. This variability indicates high values of intracluster correlation coefficient, which mean high values of design effect, and if sampled with a small number of clusters would manifest in both highly variable point estimates and very wide confidence intervals. My intuition about the outcomes comes from reviewing organ pipe plots of the key coverage indicators and seeing high variability there. Organ pipe plots are described in the 2018 WHO manual and in a 2018 Stata Conference presentation (2).

This document describes a two-fold re-examination of that data from Pakistan and Burkina Faso to, a) share the organ pipe plots of vaccination coverage outcomes, and b) explore variability and precision of estimates if those same districts were sampled using several plans, including several described in the UCLA-KSPH protocol.

# Methods

To explore the consequence of variability of coverage within a stratum on variability of the estimated coverage outcomes and their precision, we developed a simulation. Broadly speaking each run of the simulation involves these steps:

1. Define a faux population in a faux stratum with a faux outcome and exhaustive, mutually exclusive clusters, listing the % of respondents in each cluster who have the outcome of interest. The outcome can be uniform across clusters or can vary. The population may be based on data from a coverage survey or may be specified entirely *de novo* using a simple input list format. Assume that the population coverage figure is the average coverage across the population of clusters in the input dataset.
2. Plot an organ pipe plot of how the outcome is distributed across the stratum.
3. Draw repeated samples from the population, using five different sample designs, drawing 20 samples per design. Four designs are cluster samples with 5, 10, 20, and 41 clusters per draw. The sixth design is a simple random sample of 103 respondents. Four of those designs (5, 10, 41 clusters and 103 respondents) are inspired by the UCLA-KSPH document and correspondence about it.
4. From each sample, estimate the coverage point estimate and 2-sided 95% confidence interval (CI) and 2-sided 50% CI. Plot them in groups by sampling plan and indicate which CIs include the target coverage figure within the a) 95% CI and b) 50% CI.

In this document we show simulation output for several strata based on hypothetical populations and for several based on archived data from Burkina Faso and Pakistan. The slide deck that accompanies this document holds output from many strata from Burkina Faso and Pakistan.

## Details

### Defining how the outcome is distributed among the target population

The user defines how the binary outcome is distributed across the population by providing a list of equal-size PSUs and the % of persons in each PSU who have the outcome of interest.

If the outcome is distributed evenly, then every PSU will list the same coverage value. If some portions of the stratum have locally higher coverage than others, then some PSUs will list higher coverage figures and some will list lower figures. The list of input PSUs is interpreted as describing the coverage properties of the entire population of the stratum, so if the number of PSUs in the input is small, the simulation will replicate every PSU enough times to be able to draw large samples from it. Each PSU is assumed to hold the same number of potential respondents as every other PSU, and the outcome is uniformly spread within each PSU. The target coverage for the stratum is taken to be the simple arithmetic average of the PSU coverage figures in the input population.

The target population can be furnished to the simulation in two ways:

1. Using a person-level dataset with three variables: stratumid, clusterid, and a third variable holding the binary outcome coded 0 or 1
2. Using a (short or long) summary list of clusters that includes only the clusterid and the % of persons from that cluster who had the outcome of interest

The first form of input is useful for drawing samples from populations that look very much like those we have observed in survey sample datasets. The simulation can use survey data to infer how the outcome is distributed in a stratum and then simulate draws from that stratum.

The second form of output is useful for constructing simple populations of faux data with the outcome distributed however we like. The files in our GitHub repository for this work show examples of how to run the simulation with both kinds of input.

### Organ pipe plots

We summarize the distribution of the outcome across the population using an organ pipe plot. See Figure 1 for an example.

Properties of organ pipe plots:

* Every column is a cluster – in our case every column represents an equal portion of the stratum population.
* The shaded portion of the column represents persons who had the binary outcome of interest.
* The unshaded portion represents persons who did not have the outcome.
* Clusters are ordered left-to-right from highest to lowest coverage.
* The shaded portion of the plot is equal to the estimated coverage for that outcome. In figure 1, 60.0% of the figure is shaded and 60.0% of respondents in that stratum have the outcome of interest, but the outcome is more common in some parts of the stratum and less in others.
* Our plots show the population coverage and the intracluster correlation coefficient (ICC)
* For upcoming work where we consider LQAS sampling, the standard deviation of the outcome across clusters will be important, so we have noted it here in a footnote.

Chart, histogram

Description automatically generated

Figure . Example of an organ pipe plot to summarize how a binary outcome is distributed across portions of a stratum

### Simulation sampling plans

Each run of the simulation executes each five sampling plans 20 times. The properties of the sampling plans are set inside the program – they are the same for all the work we describe in this document. The 100 confidence intervals that result from each run are portrayed in a single plot that complements the organ pipe plot. See Figure 2.

From top to bottom in Figure 2, the sampling plans are:

1. A simple random sample of N=103 respondents – inspired by the UCLA-KSPH doc
2. A sample of 41 PSUs with *m*=10 respondents per PSU – from the UCLA-KSPH doc
3. A sample of 20 PSUs with *m*=4 respondents per PSU
4. A sample of 10 PSUs with *m*=8 respondents per PSU
5. A sample of 5 PSUs with *m*=32 respondents per PSU – from the UCLS-KSPH doc

The cluster samples draw a bootstrap sample of the target number of clusters, and then within each cluster, draws a simple random sample of *m* respondents. Then each sample is used to estimate coverage and calculate two 2-sided survey adjusted logit confidence intervals (CIs): one to show the region of 95% confidence and another shorter one to show the region of 50% confidence (3).

Within each sampling plan, the 95% confidence intervals are plotted in a stack which is sorted by the lower bound of the 95% CI. The target coverage figure (average cluster coverage in the population dataset) is shown with a continuous vertical red line.

The point estimates are colored using this logic:

1. If the sample design effect is <= 1.5, the point estimate is shown in **black**.
2. If the sample design effect is > 1.5, the point estimate is shown in **red**.

The confidence intervals are colored using this logic:

1. If the 95% CI does not include the target coverage figure, the entire CI is **red**.
2. If the 95% CI includes the target figure, the 95% CI is **gray**.
3. If the 50% CI includes the target figure, the 50% CI is **green**.

Within each sampling plan we expect, on average, one CI to be colored red, and ten CIs to have centers that are colored green.

The properties of the 20 samples are summarized in tabular form at the top of each stack of intervals. A row of text lists the sample size, the number of respondents per PSU (*m*). It also lists the median and range (min-max) of the 20 observed values of: estimated coverage (Cvg), design effect (DEFF), effective sample size (NEFF), and 2-sided 95% CI half-width[[1]](#footnote-1) (HW).

A picture containing diagram

Description automatically generated  
Figure . Results from a single simulation run -- six sets of 20 2-sided 95% confidence intervals (CIs).

Each confidence interval plot is combined with the corresponding organ pipe plot into a single plot file. See Figure 3.

# Results

Figure 3 shows a stratum from a recent survey in Pakistan (TPVICS Round 2, reports forthcoming) with an outcome whose coverage varies tremendously, with nearly all the respondents in some portions having the outcome, while only a very few respondents in other portions have the outcome. The ICC is 0.208 or about 1/5. In this case, the CIs for samples with 5 PSUs have very wide confidence intervals and their point estimates vary from 39-81% and the design effect varies from 2.5 to 16.4 with a median of 6.9. Some of the sample point estimates fall more than 15 percentage points away from the target figure. As the number of PSUs increases, moving up the figure, the confidence intervals get more narrow as do the range of estimated outcomes and the design effects. The design effect for every cluster sample shown here exceeds 1.5, which was a target figure mentioned in the UCLA-KSPH design document for the design with 5 PSUs.

Figure 4 includes many PSUs where some respondents have the outcome, but also numerous PSUs where the outcome is entirely absent. The ICC is 0.27 or just over 1/4. Again, the outcomes with 5 PSUs are quite wide with half-widths as wide as 37% and they become more narrow as the number of PSUs increase.

Figure 5 draws from a fictional stratum where the outcome is uniformly distributed; in every portion of the stratum, 75% of eligible population members have the outcome. In this case even the sampling plan with 5 PSUs yields narrow confidence intervals.

Figure 6 is another stratum from Pakistan. Its outcome is distributed in a fairly uniform fashion. The samples with 5 PSUs here are more narrow than in some of the earlier examples.

There is a PowerPoint slide deck that accompanies this document that holds simulation outputs for two outcomes: % of children fully-vaccinated and % of children who are zero dose, for 152 districts in Pakistan from 2022 and 63 districts from Burkina Faso in 2016. Review them at your convenience.

Chart

Description automatically generated

Figure . Simulation results when the cluster level coverage varies widely. Stratum 315 from Pakistan TPVICS Round 2 Survey, 2022

Chart, histogram

Description automatically generated

Figure . Simulation results when the outcome is entirely absent from some portions of the stratum. Stratum 123 from Pakistan TPVICS Round 2 Survey, 2022

Graphical user interface, application, table, Excel

Description automatically generated

Figure . Simulation outcome from a fictional stratum with uniform 75% coverage

Chart

Description automatically generated

Figure . Simulation output from a stratum with nearly uniform coverage. Stratum 471 from Pakistan TPVICS Round 2 Survey, 2022

# Discussion

The simulation output shows that when the underlying outcome has heterogenous coverage within the stratum, samples with only five PSUs yield extremely wide confidence intervals and yield point estimates that sometimes fall very far from the true population figure. Their design effects are nearly always higher than 1.5. In many settings where a coverage survey is conducted, it will simply not be true that all of the outcomes of interest are distributed uniformly within every stratum of interest, so the five PSU design will not be a credible all-purpose alternative to the WHO 2018 design that recommends 30 or more PSUs.

Within any specific stratum, some outcomes may be distributed uniformly across the stratum while others may vary. Different outcomes in the same survey will have different values of estimated coverage and different values of ICC in the same stratum, and therefore different markedly confidence interval widths.

The figures for district level outcomes in Pakistan and Burkina Faso show little evidence of uniform coverage except for outcomes that are nearly 100% or nearly 0%. When stratum level coverage is between say 20-80%, we see quite a lot of variability in cluster level coverage, which means that the samples with five and 10 PSUs have notably wider confidence intervals and notably more spread in their point estimates than those with more PSUs.

The UCLA-KSPH team suggests that in DRC and possibly CAR the outcomes of interest might be distributed in a uniform fashion at the district level. That hypothesis will be nicely tested with their plans to use 41 PSUs with ten target respondents each in some of the upcoming work.

Finally, I note that one or more recent surveys in DRC has used five PSUs per district, and based on these simulations, I recommend that:

1. It would be helpful to report confidence intervals along with point estimates, because some of those intervals are likely to be very wide.
2. It would be important to take the sample design into account when calculating those confidence intervals.
3. Even if the underly coverage outcome did not change from year to year, in consecutive samples, the outcome estimates for a fixed district could easily differ by more than 10 or even 20 percentage points. We see some differences that are much larger even than 20% in the five and 10 PSU samples in these simulations. So it would be very challenging to interpret differences from survey rounds and to identify what portion of those differences are due to sampling variability and what portion might be due to changes in the underlying outcome over time. In order to confidently interpret trends in coverage over time, it is important to design for narrow confidence intervals.

# Next Steps

I would be happy to go over this material and address any questions in a teleconference sometime soon at a time of day when we can have participants from both KSPH and UCLA. You can use this link to find a time on my calendar: <https://calendly.com/dale-rhoda/60min>.

The Stata programs described here are available for download from the GitHub repository at: <https://github.com/BiostatGlobalConsulting/bgc-ucla-ksph-sampling-simulation>. You may use the files there to run this simulation on data from the recent DRC survey or other surveys or on data from faux hypothetical distributions of the outcome.

Several of those programs (svypd.ado, opplot.ado, and their associated .sthlp files) are also included in our Vaccination Coverage Quality Indicators (VCQI) collection of programs that were developed for the World Health Organization (4).

I’m happy to help orient you to using the software.

In the next week or two, I will either write a second simulation or adapt this first one to also include the LQAS-based sampling variation described in the UCLA/KSPH study design document. I will share output from that program when it is available.

# References

1. World Health Organization. Vaccination Coverage Cluster Surveys: Reference Manual [Internet]. March 2021 (WHO/IVB/18.09). License: CC BY-NC-SA 3.0 IGO. Geneva, Switzerland: World Health Organization; 2018 [cited 2021 Jul 27]. Available from: https://apps.who.int/iris/handle/10665/272820

2. Prier ML, Rhoda DA. Organ Pipe Plots for Clustered Datasets - Visualize Disparities in Cluster-Level Coverage [Internet]. Stata Conference 2018; 2018 Jul 19 [cited 2020 May 6]; Columbus, Ohio. Available from: https://github.com/BiostatGlobalConsulting/organ-pipe-plots/blob/master/opplot\_presentation.pptx

3. Dean N, Pagano M. Evaluating Confidence Interval Methods for Binomial Proportions in Clustered Surveys. J Surv Stat Methodol. 2015 Dec;3(4):484–503.

4. Biostat Global Consulting. Vaccination Coverage Quality Indicators (VCQI) Resources [Internet]. 2017. Available from: http://www.biostatglobal.com/VCQI\_resources.html



+1 (614) 499-2351

Dale.Rhoda@biostatglobal.com

www.biostatglobal.com

1. These confidence intervals are asymmetric when the coverage estimate is near 0% or 100%, so in other contexts, sometimes people report the length of the longer of the two CI sides as the half-width. That is not what we mean here. In this document, CI half-width is literally half the distance from the lower to upper bound of the 2-sided 95% CI. [↑](#footnote-ref-1)