Afghanistan Serosurvey Supplement

Report generated on: 2018-04-23

This document was created by RMarkdown in Rstudio Version 1.0.143.

Source file: polio\_merged7.sas7bdat

# Background

This serosurvey used existing specimens from a nutrition survey conducted in Afghanistan during 2013 as a collaborative effort between Afghanistan Ministry of Public Health and Agha Khan University. The nutrition survey employed a stratified two-stage sample design with 30 enumeration areas (EAs) selected from each of 34 provinces. A micronutrient assessment collected serum specimens from participants in a subset of 80 EAs from 26 provinces. There were 981 participants in the micronutrient assessment, of which 690 (70%) were available for testing with a polio neutralization assay

This supplement documents the multiple imputation process applied using the mi package.

**Survey design variables**:

\*stratum (stratum)

\*first stage cluster (psu)

\*sampling weight adjusted for non-response in the original nutrition survey (samp.wts)

**Geographic variables**:

\*5 regions:central, eastern, north, south, west (region)

**Demographic variables**:

\*age of child (agemos)

\*child's sex (sex)

\*mother's education: Illierate, PreSchool/Religious/Some Schooling, High School/14+years schooling/Bacchelors Degree (momeduc)

*NOTE: 43% of the 981 observations are missing mothers education info (13% of observations are missing both mother's education and outcome variables (Sabin1,Sabin2, Sabin3))*

**Health/micro-nutrition**:

\*height

\*weight

\*diarrhea in the past 2 weeks (yes/no) (diarr)

\*hemoglobin (g/dL) (hb)

*NOTE: 11% of observations are missing child's height (4% of observations are missing both weight and Sabin1)*

*NOTE: 11% of observations are missing child's weight (4% of observations are missing both weight and Sabin1)*

*NOTE: 42% of observations are missing diarrhea variable (13% of observations are missing both diarrhea and Sabin1)*

*NOTE: 7% of observations are missing hemoglobin variable (0.5% of observations are missing both hemoglobin and Sabin1)*

**Primary outcomes of interest**:

\*neutralizing antibody titer for sabin type1 (Sabin1)

\*neutralizing antibody titer for sabin type2 (Sabin2)

\*neutralizing antibody titer for sabin type3 (Sabin3)

*NOTE: either all 3 polio titers outcomes are observed or all are missing*

*NOTE: For the final analysis titers are dichotomitized into seropositive for titers* $\geq $ *8, or seronegative for titer < 8.*

# Missing Data Patterns



Figure 1: Missing Data Patterns



Figure 2: Organ pipe plot of % observed Sabin titers by region. Sample size is given at base of each bar.

## Margin Plots

\*Blue dots: both observed

\*Red dots on horizontal access: x-axis variable observed, y-axis variable missing

\*Red dots on vertical access: y-axis variable observed, x-axis variable missing

\*Yellow in lower left corner: some observations are missing both



Figure 3a: Margin plot of Sabin1 titer and age in months.



Figure 3b: Margin plot of Sabin1 titer and hemoglobin.



Figure 3c: Margin plot of Sabin1 titer and child's weight.

|  |  |  |
| --- | --- | --- |
| sex | Sabin1 | Count |
| Male | Neg | 5 |
| Male | Pos | 346 |
| Male | NA | 157 |
| Female | Neg | 13 |
| Female | Pos | 326 |
| Female | NA | 134 |

Table 1: sex by pos/neg for Sabin1.

|  |  |  |
| --- | --- | --- |
| momeduc | Sabin1 | Count |
| Illit/ | Neg | 7 |
| Illit/ | Pos | 299 |
| Illit/ | NA | 118 |
| Pre/Relig/Some | Neg | 3 |
| Pre/Relig/Some | Pos | 69 |
| Pre/Relig/Some | NA | 31 |
| High,14yrs,Bach | Neg | 1 |
| High,14yrs,Bach | Pos | 21 |
| High,14yrs,Bach | NA | 10 |
| NA | Neg | 7 |
| NA | Pos | 283 |
| NA | NA | 132 |

Table 2: Mom's Education by pos/neg for Sabin1.

|  |  |  |
| --- | --- | --- |
| Diarrhea | Sabin1 | Count |
| Yes | Neg | 4 |
| Yes | Pos | 132 |
| Yes | NA | 75 |
| No | Neg | 7 |
| No | Pos | 266 |
| No | NA | 87 |
| NA | Neg | 7 |
| NA | Pos | 274 |
| NA | NA | 129 |

Table 3: Diarrhea by pos/neg for Sabin1.

# Imputation

Imputation was completed using the mi package version 1.0. Details can be found at [https://CRAN.R-project.org/package=mi](https://CRAN.R-project.org/package%3Dmi), including the package documentation and a vignette by Goodrich et al. In addition an article Su, et al. [2].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | type | missing | method | model |
| region | unordered-categorical | 0 |  |  |
| Province | unordered-categorical | 0 |  |  |
| stratum | unordered-categorical | 0 |  |  |
| psu | unordered-categorical | 0 |  |  |
| samp.wts | continuous | 0 |  |  |
| agemos | positive-continuous | 0 |  |  |
| sex | binary | 0 |  |  |
| momeduc | ordered-categorical | 422 | ppd | ologit |
| hb | positive-continuous | 72 | ppd | linear |
| diarr | binary | 410 | ppd | logit |
| height | positive-continuous | 111 | ppd | linear |
| weight | continuous | 112 | ppd | linear |
| Sabin3 | bounded-continuous | 291 | ppd | linear |
| Sabin2 | bounded-continuous | 291 | ppd | linear |
| Sabin1 | bounded-continuous | 291 | ppd | linear |

Table 4: Variable list with number missing values and the variable types

In choosing a variable type for imputation of the Sabin variables, we considered continuous, positive-continuous, bounded continuous, and ordered categorical. We opted for bounded continuous though we saw little difference when using ordered categorical (28-30 unique values).

\*R Syntax:

##########################################################
# Imputation using the mi package

library(mi)

mdf <- missing\_data.frame(as.data.frame(temp2))

show(mdf)

mdf <- change(mdf, y = c("stratum", "momeduc","agemos","hb","weight","height"), what = "type",
 to = c("un", "or",rep("positive-continuous",4)))

mdf <- change(mdf, y = c("Sabin1","Sabin2","Sabin3"), what = "type",
 to = c(rep("bounded-continuous",3)), lower=5, upper=1449)

show(mdf)

> imp <- mi(mdf, n.iter = 30, n.chains = 10,max.minutes = 300)
> Rhats(imp)

# 15 additional iterations
> imp <- mi(imp, n.iter = 15,max.minutes = 300)
> Rhats(imp)

#plot(imp)

#summary(imp)

# Imputation diagnostics

The following diagnostic plots are derived from the plot function in the mi package.



Figure 4: Diagnostic plots of first 3 chains of imputation.

|  |  |
| --- | --- |
| mean\_momeduc | 1.0130076 |
| mean\_hb | 1.0412961 |
| mean\_diarr | 1.0203095 |
| mean\_height | 1.0049738 |
| mean\_weight | 1.0187661 |
| mean\_Sabin3 | 0.9946073 |
| mean\_Sabin2 | 1.0001075 |
| mean\_Sabin1 | 0.9995673 |
| sd\_momeduc | 1.0005806 |
| sd\_hb | 1.0671859 |
| sd\_diarr | 1.0074987 |
| sd\_height | 1.0287234 |
| sd\_weight | 1.0382263 |
| sd\_Sabin3 | 0.9906683 |
| sd\_Sabin2 | 0.9961686 |
| sd\_Sabin1 | 0.9937356 |

Table 5: R statistic (if <1.1 then the imputation is considered coverged).

# Data preparation

Finally we extracted 30 imputed datasets. Thirty was chosen because ~30% of the primary outcome observations of interest, Sabin titer for type 1, 2 and 3, were missing.

# Data analysis

The analysis of each imputed dataset was completed in the survey package. We used the adjusted degrees of freedom as described on page 358 from the book: Applied Survey Data Analysis, Heeringa, West and Beglund (2010).

# R Packages

\*VIM

Alexander Kowarik, Matthias Templ (2016). Imputation with the R Package VIM. Journal of Statistical Software, 74(7), 1-16. doi:10.18637/jss.v074.i07

\*MI

Gelman A and Hill J (2011). “Opening Windows to the Black Box.” *Journal of Statistical Software*, *40*.

Vignette by Ben Goodrich and Jonathan Kropko, for this version, based on earlier versions written by Yu-Sung Su, Masanao Yajima, Maria Grazia Pittau, Jennifer Hill, and Andrew Gelman. "An example of mi Usage"

\*ggplot2

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

* survey T. Lumley (2017) "survey: analysis of complex survey samples". R package version 3.32.
* mitools

Thomas Lumley (2014). mitools: Tools for multiple imputation of missing data. R package version 2.3. [https://CRAN.R-project.org/package=mitools](https://CRAN.R-project.org/package%3Dmitools)