I am currently completing a project that calculates household and women's phone ownership estimates by demographic characteristics (wealth index, gender, rural/urban etc.) across several countries' DHS surveys.

While running the standard code in R for calculating women's phone ownership (I provide the code/function below), I noticed that there is a high rate of NAs in the India data, higher than any other country's (in my dataset spanning 43 country-year surveys). If I run the standard code, I get a low phone ownership rate of approximately 8%. If I omit all NAs (84% of the data!), I still get phone ownership = 53.9% among women, which is higher than the rate stated in the official DHS report and in StatCompiler (50.9%)

Is there a reason for this?

Function for calculating women's phone ownership:

```
analyze_dhs_wm <- function(df_name) {
  df <- get(df_name, envir = .GlobalEnv)
  country_code <- substr(df_name, 1, 3)

  if (!"v169a" %in% names(df)) {
    # Handle the case where v169a does not exist in the dataframe
    phone_ownership_weighted <- NA
    phone_ownership_percent <- NA
    na_percentage <- NA
  } else {
    # Proceed if v169a exists
    phone_ownership_weighted <- sum(df$v169a * (df$v005 / 1e6), na.rm = TRUE) / sum(df$v005 / 1e6, na.rm = TRUE)
    phone_ownership_percent <- phone_ownership_weighted * 100
    na_percentage <- mean(is.na(df$v169a)) * 100
  }

  return(data.frame(country_code = country_code,
          type = "DHS_wm",
          phone_ownership_weighted = phone_ownership_weighted,
          phone_ownership_percent = phone_ownership_percent,
          na_percentage = na_percentage))
}

results_dhs_wm <- do.call(rbind, lapply(dhs_wm_data_frames, analyze_dhs_wm))
```

Following is a response from Senior DHS staff members, Tom Pullum and Fred Arnold:

First, the NA responses are due to subsampling. In NFHS-5, Question 931 for women ("Do you have any mobile phone that you yourself use?") and Question 932 for women ("Do you use your mobile phone for any financial transaction?") are only asked of women with the household selected for the state module. If the household is not selected for the state module, questionnaire sections 8 through 11 are skipped. This is described in Section 1.2 of the NFHS-5 final report (see below):

"NFHS-5 was designed to provide information on sexual behaviour; husband's background and women's work;
HIV/AIDS knowledge, attitudes, and behaviour; and domestic violence only at the state level (in the state module),
while indicators in the district module are reported at the district level. A subsample of 15 percent of households was
selected for the implementation of the state module drawn from the district sample. In 15 percent of households
randomly selected for the state module, a long questionnaire was administered that included all the questions needed
for district-level estimates plus additional questions for the topics listed above. To achieve a representative sample
of 15 percent of households, NFHS-5 conducted interviews in every alternate selected household in 30 percent of the
randomly selected clusters."

Second, in table 14.23 of the NFHS-5 report, on page 637, top row, the percentage is given as 53.9%, the percentage you got with R. In Stata, we get the same thing:

. tab v169a if v169a<. [iweight=v005/1000000]

```
    owns a |
    mobile |
 telephone |     Freq.    Percent     Cum.
------------+-----------------------------------
        no | 49,744.866    46.05     46.05
       yes | 58,268.694    53.95    100.00
------------+-----------------------------------
     Total | 108,013.56    100.00
```

(The 53.95 in this table is actually 53.94572%, which rounds to 53.9%.) If STATcompiler has something different, then that would be inconsistent, but what you got seems to be ok.

Thank you so much! This is super helpful.