

1. I am going to work on my Study 2 alone.

2. Proposal summary for Study 2

- Title: The effect of family wealth on the practice of female genital mutilation/cutting in the next generation of Sudanese.
- Key research question: Is there an impact of the level of wealth on the household size, the level of women's education and the prevalence of female genital mutilation/cutting in Sudan?

Sudan has been known to have one of the world's highest rates of female genital mutilation/cutting (FGM/C). The practice of FGM/C, or female circumcision, has been prevalent in certain countries for a non-medical reasons, only threatening women's health. It causes complications during their menstruation, pregnancy, delivery, and post-delivery, in other words, throughout their lives. Furthermore, these diseases often result in exclusion of the women from their families and societies. In the communities that practice it, it is thought to ensure virginity of women before marriage and their fidelity to their husbands afterward, and to enhance sexual pleasure on the males' side. It raises global concerns as in the countries where the practice is common, the populations usually increase more rapidly compared to other regions; it also means there are larger and larger populations at risk of getting exposed to this unhealthy practice in the immediate future.

I have felt terrible every time I heard about this topic, and become greatly interested in the community of Sudan itself as FGM/C is known to be largely dependent on the cultural factors of the families and local communities. I would like to see if the wealth level of households has any association with (1) the household size which is directly related to the total population size,

(2) the education level of mothers, and (3) the circumcision proportion in the mothers and daughters. My hypothesis is the wealth score of households has a direct relationship with the education level in mothers, and an inverse relationship with the household sizes and the proportion of FGM/C in mothers and daughters in Sudan.

3. Data set description

This data set shows the results of Multi Indicator Cluster Surveys (MICS) round 5 in Sudan conducted by UNICEF in 2014. The MICS has been conducted since September, 1991 in 108 countries so far, and its round 6 is currently (October, 2017) ongoing in 43 countries.

In this survey in Sudan, three different questionnaires were used for interviews—about (1) household, (2) women aged 15 to 49, and (3) children under five. The household listing and interviewer training were done in July, 2014, and survey fieldwork was carried out from September 10 to September 14, 2014. 18,000 households were sampled, and 16,801 responded to interviews with a response rate of 98%. 20,327 women were eligible for interviews, and 18,302 responded to the interviews with a response rate of 90%. For “children under five” questionnaire, mothers/caretakers were interviewed with the response rate 95.5%. I am interested in a dataset named “fg”, which is a combination of parts from household and women questionnaires, and a dataset “hh” which includes the result of household questionnaires. Referring to the final report of UNICEF on this survey, the percentage of women aged 15-49 who stated that “FGM/C should be continued” was reported as 40.9. The percentage of women aged 15-49 who reported to have undergone any form of FGM/C was 86.6, and percentage of daughters aged 0-14 who had undergone any form of FGM/C as reported by mothers was 31.5.

The subjects under the questionnaires relevant to the two datasets were the households with

an average size of 5.9 and women aged 15 to 49 in Sudan. The original dataset of “fg” includes 21,947 observations on 24 variables. After I tidied the table in R so it only included the variables wscore, HH2, FG9, FG3, FG15, and welevel, 17,613 observations remained in the table. After removing inputs “9” which indicates “don’t know” in this dataset, 17,542 rows remained. The dataset “hh” includes 12,150 observations on 167 variables. I am only going to use the column HH11 from the table. The range of the values for column HH11 (the number of household members) is between 1 and 26. I am planning to left-join the tables fg and hh, having the household number (HH2) as the common variable for the joining process. The access to the dataset was granted by UNICEF at my request, and the data were provided as a SAV file.

My quantitative outcome is going to be the combined wealth score (wscore). I have learned on the MICS website that it is a measurement of the wealth level of households which combines the urban wealth score, a measure obtained only in the households in urban areas, with the rural wealth score, observed only in rural areas. Thus, combined wealth score is a measure of the overall national level wealth index for the household. Its range in the dataset is approximately between -1.4 and 3.4. The higher it is, the wealthier the household is. My predictors for the outcome are the number of household members (HH11), the only variable from the table hh, the number of daughters (FG9), a binary variable on whether the mother has ever been circumcised (FG3), a binary variable on whether the daughter has been circumcised (FG15), and the categorical variable on the education level of the mother (welevel).

The codebook is part of the final report by UNICEF, a pdf link (page 319, 371-373):

https://mics-surveys-prod.s3.amazonaws.com/MICS5/Middle%20East%20and%20North%20Africa/Sudan/2014/Final/Sudan%202014%20MICS_English.pdf