**Exploring linkages between Domestic Violence   
& Infant Mortality in India**

Anushree Deb

10th May 2019

LBJ School of Public Affairs

University of Texas Austin

**Appendix: Data Management Plan**

### dataset

The data used for the purpose of this study is the DHS India 2015 -16 data (<https://www.dhsprogram.com/>). Known as The Demographic and Health Survey, it is administered by the DHS Program, funded by a combination of various donor agencies like the USAID, Bill and Melinda Gates Foundation, DFID, UNICEF, Mac Arthur Foundation etc. DHS coordinates and liaisons with several government departments and helps conduct locally administered health and demographic surveys.

The primary objective of the survey is to collect health and family welfare data through clinical, anthropometric, & biochemical questions. The surveys are administered through household, woman’s, man’s & biomarker questionnaire. In order to maintain comparisons and similarity across data, DHS tries to ensure standardized questions irrespective of regions. However, area and country specific questions are allowed provided they receive appropriate institutional approval.

In India, the DHS Surveys are administered under the Ministry of Health and Family Welfare, coordinated by the International Institute for Population Sciences, Mumbai, and implemented by groups of survey organizations. “*The NFHS is a collaborative project of the International Institute for Population Sciences (IIPS), Mumbai, India; ORC Macro, Calverton, Maryland, USA and the East-West Center, Honolulu, Hawaii, USA. The Ministry of Health and Family Welfare (MOHFW), Government of India, designated IIPS as the nodal agency, responsible for providing coordination and technical guidance for the NFHS”.[[1]](#footnote-1)* The surveys collect information on household population, health, nutrition, and welfare-related questions across 29 states and 7 union territories in India. Though their sampling strategy is quite extensive, they sample 157 districts that have a similar urban/rural split of 30/70 like the rest of the country and then use proportional representation to calculate further estimates.

Since the study required the generation of several independent and dependent variables, a specific set of variables have been used from the original dataset to compute them and conduct the required analysis. The original dataset has over 1300 variables and it was not possible to include them all since the range of information was too diverse, ranging from HIV status to current blood pressure.

### Data Sharing & Format

The data type used is STATA files in a. dta format. Sharing and long-term access to the data are available through the use of STATA. Individual or group users have to submit a research abstract to DHS which takes 2-3 days to approve the request. Upon approval, researchers are able to download the data they need. Once downloaded, the data is not meant to be shared without further approval and should remain with the primary researcher itself. To facilitate teamwork, multiple users can submit a joint proposal and access the data together. This eliminates sharing and updating concerns. Since the file size is too big to upload on GitHub, birth\_recode.dta has been uploaded on google drive and can be accessed through the following link:

[**https://drive.google.com/open?id=1pj6io71D3k4gKKZE0y-YJ6Bmb3Yg3V-H**](https://drive.google.com/open?id=1pj6io71D3k4gKKZE0y-YJ6Bmb3Yg3V-H)

### Ethics and Legal Compliance

Since this data is being used for academic research purposes only, there will be no ethical and compliance issues. Additionally, all of this data has been anonymized and there is no scope for misuse.

**IPR**

DHS Data has been cited and sourced at all relevant times and DHS owns IPR to the data

### Data Storage & Nomenclature

In order to ensure timely backups and no missing copies, I have saved 2 copies of the data in separate locations. One under Meta Data, which holds all original data files and remains untouched. This is for backup and emergency purposes only. Another copy of the file is saved under the assignments folder and this version is used to conduct data analysis. Since DHS is survey data, it does not need to be updated. All files were named according to the

**For Data Analysis**

iCloud Drive▸ Desktop ▸ MGPS ▸ Data Management ▸Assignments▸ Final Paper ▸ Data ▸ DHS India 2015-16▸ DHS\_Birth Recode

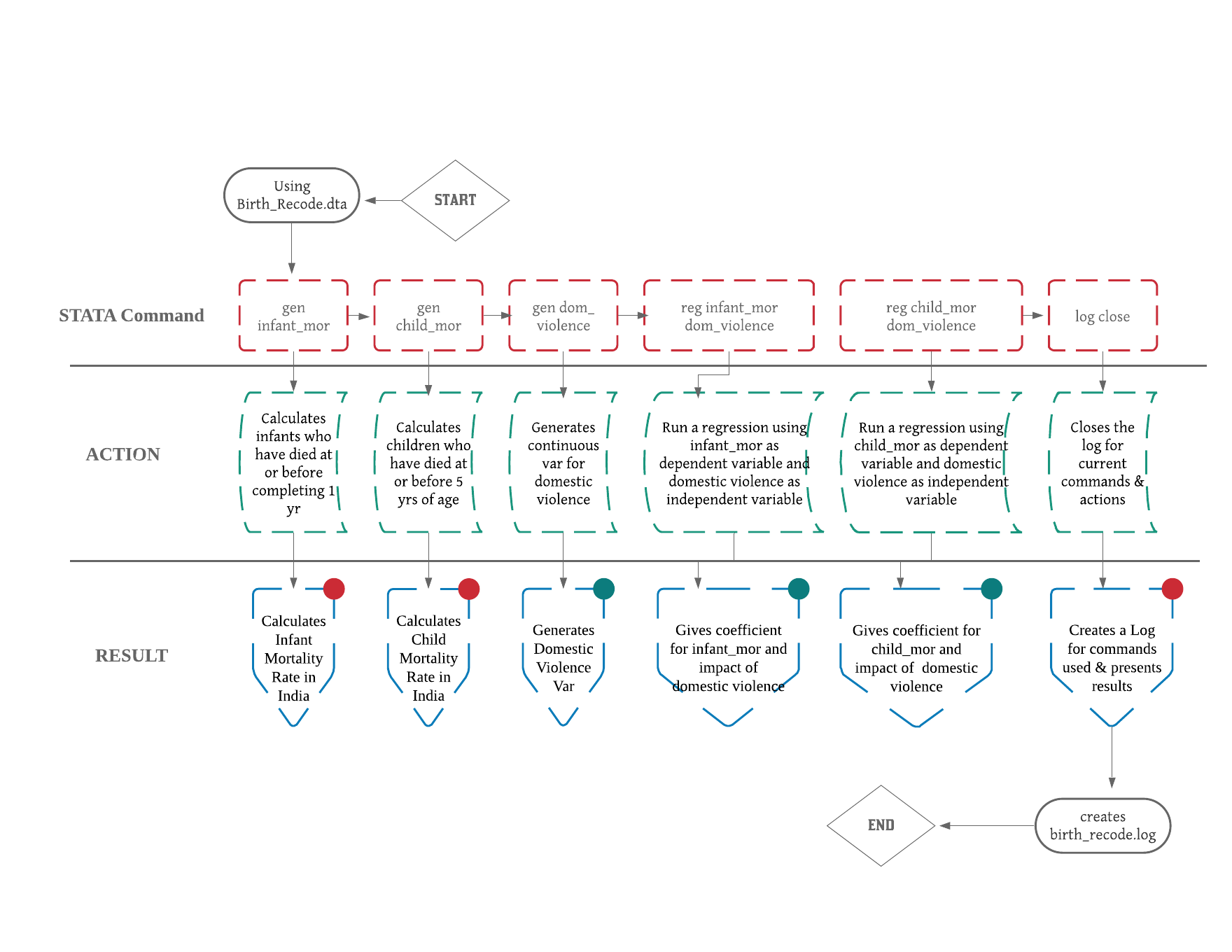
**For Backup**

iCloud Drive▸ Desktop▸ MGPS▸ Data Management ▸Meta Data ▸DHS India 2015 -16 ▸ DHS\_Birth Recode

For ease of use, the .dta file for DHS Birth\_Recode was renamed to birth\_recode.dta. The do and log file for the project were renamed to IMR and birth\_recode.log respectively

### Selection and Preservation

DHS Birth\_Recode is the most important dataset within the data and should be retained until the end of the Spring Semester when the course is finished, and the assignment has been graded. This will be done to ensure there is no discrepancy between the assignment submitted and the actual findings.

**WORKFLOW**

**Figure 3: Workflow Operationalization**

### Variable List

1. Dependent variables
   1. Infant Mortality Rate (IMR)
   2. Child Mortality Rate (CMR)
2. Independent variable
   1. Dom\_Violence: includes a continuous index for Domestic Violence. Missing values are converted and then dropped from the variable. It includes yes, no and don’t know answers only.
   2. Wealth\_index: a continuous index generated for households measuring wealth. Households are given scores based on the number and kinds of consumer goods they own, ranging from a television to a bicycle or car, and housing characteristics such as source of drinking water, toilet facilities, and flooring materials. These scores are derived using principal component analysis.
   3. Edu: Educational Attainment of the Mother in single years
   4. Height: Height of the mother
   5. Urban/Rural: dummy variable which indicates of the location if rural or not

### OPERATIONALIZATION DEFINITIONS

1. **Infant Mortality Rate**: the number of infant deaths (died when age was less than or equal to 12 months) in a given year by 1,000 live births in the same year.
2. **Child Mortality Rate**: the probability per 1,000 live births that a newborn baby will die before reaching age five
3. **Domestic Violence**: Domestic violence is a pattern of behaviors used by one partner to maintain power and control over another partner in an intimate relationship. It includes behavior that causes physical harm, arouse fear or force them to behave in ways they do not want. It includes the use of physical and sexual violence, threats and intimidation, emotional abuse and economic deprivation.[[2]](#endnote-1)

**CODEBOOK & DATA DICTIONARY**

***For computing IMR & CMR***

1. V008: Date of Interview (cmc)
2. B3: date of birth (cmc)
3. b7: age at death (months, imputed)
4. CMC: Century-Month Code system is how dates are coded in Demographic and Health Surveys. CMC reduces all months to a code taking the value of 1 in January 1900, 2 in February 1900, 13 in January 1901 etc.[[3]](#footnote-2)
5. Age: is calculated by subtracting date of the interview from the date of birth to calculate the age of the child
6. Child\_age: includes all age and missing values for age of children
7. Infant\_mor: number of infants who have died when their age has been less than 12 months
8. Child\_mor: number of children who have died when their age has been less than 5 years
9. 0 and 1000 are used in infant\_mor and child\_mor when 0 means the infant/child is alive and 1000 means when the infant/child is not

***For computing Domestic Violence***

1. V744a: Beating justified if wife goes out without telling husband
2. V744b: Beating justified if wife neglects the children
3. V744c: Beating justified if wife argues with husband
4. V744d: Beating justified if wife refuses to have sex with husband
5. V744e: Beating justified if wife doesn't cook food properly
6. S936f: Justifies domestic violence: Wife unfaithful
7. S936g: Justifies domestic violence: Wife disrespect

**REPLICABILITY**

In order to achieve replicability for the study, the code for the entire study is stored in the STATA do file labelled “IMR”. Running the code from the beginning will replicate the study and generate the exact same results as have been depicted here. A log file named “birth\_recode.log” which contains all the results generated from the code in a sequential manner has also been provided. The log file replicates itself each time the code is run and stores the updated results automatically. Both of these files along with the paper and the appendix have been uploaded in the GitHub folder “Final Paper”.

1. <http://rchiips.org/nfhs/about.shtml> [↑](#footnote-ref-1)
2. <https://www.thehotline.org/is-this-abuse/abuse-defined/> [↑](#endnote-ref-1)
3. <http://demographicestimation.iussp.org/content/dhs-century-month-codes> [↑](#footnote-ref-2)