

Pipeline Overview and Data Source

1 Summary

In this document, we give an overview of the pipeline and suggest the order at which scripts are run. In addition, we summarize the main data source for U5MR pipeline, specifically on shapefile and DHS data. For preparation of population rasters, please reference the vignette ‘Urban Fraction for Stratified Beta-binomial Model’.

2 Pipeline Overview

The pipeline is structured as three parts:

1. Obtaining direct and smoothed direct estimates. They are standard ways of estimating U5MR at national or Admin-1 level, but they could hardly compute Admin-2 level estimates for the reasons described in the main report. They serve as comparisons for our beta-binomial model estimates. The scripts are under ‘Rcode/Direct_Smooth’.
2. Stratified beta-binomial model is our main approach for analysis. The scripts are under ‘Rcode/UR Pipeline’. We first calculate urban fractions for under 5 population, then fit stratified model and finally get the overall aggregated estimates. More details could found in another vignette, ‘Urban Fraction for Stratified Beta-binomial Model’.
3. Scripts that produce maps and figures are under ‘Rcode/Visualization’. We provide scripts for the set of figures shown in the main report and appendix.

The detailed instruction about implementing the scripts in order is:

1. Create a new directory and put the entire folder “Rcode” and the script named “create_folder.R” into the directory.
2. Run “create_folder.R”. Make sure to specify the country. This creates the folder structure to save data and results.
3. Gather the source data described in this document and put them into the folder named “Data”.
4. Run “Rcode/dhs_prepare.R”. Make sure to specify the years and file names.
5. Run “Rcode/Direct_Smooth/Direct_SmoothDirect.R” to obtain direct and smoothed direct estimates.
6. Run scripts under the folder “Rcode/UR Pipeline”. First run “prepare_thresh.R”, then “thresh.R” and finally “stratified_BB8.R”. The first two scripts prepares all required data for stratified Beta-binomial model and the last one fit the models and generate estimates. Please reference another vignette, ‘Urban Fraction for Stratified Beta-binomial Model’, for getting population data and detailed description.
7. To replicate the plots in the DHS report, run scripts in the folder named “Rcode/Visualization”. “3UR_ratio_report.R” produces Figure X.4 in the appendix of the main report. “compare_report.R” creates Figure X.2 in the appendix of the main report. “report_main.R” generates the rest of figures appeared in the appendix. All figures are stored under “Results/country/Figures/Report”.

3 DHS Survey Data

We need the BR (birth) recode for the DHS survey for U5MR estimation. To obtain the data, one needs to register first at the DHS website (<https://dhsprogram.com/>). After submitting a data request with description for the project, the data will be available to the user within days.

Both the survey data and associated GPS data serve as input for SUMMER preprocessing scripts. The user could navigate to the interface below and proceed to data download. Here we use Zambia as an example.

Download Approved Datasets or Request Additional Datasets for Existing Project(s)

Download by Single Survey
 Spatial analysis of childhood mortality and fertility rate in Africa
 (*) denotes restricted datasets
 Select Another Country

Zambia ▼

Please click on the "Download" link to download datasets for a specific survey or click the "Country/Year" link to view the survey information page.

Country/Year	Type	Phase	Survey Datasets	GPS Datasets	HIV/Other Biomarkers Datasets	SPA Datasets
Zambia 2018	Standard DHS	DHS-VII	Download	Download		
Zambia 2013-14	Standard DHS	DHS-VI	Download	Download		
Zambia 2007	Standard DHS	DHS-V	Download	Download		
Zambia 2005	HIV SPA	DHS-V				
Zambia 2001-02	Standard DHS	DHS-IV	Download			
Zambia 1996	Standard DHS	DHS-III	Download			
Zambia 1992	Standard DHS	DHS-II	Download			

Figure 1: Downloading page on DHS website

The BR recode data should be downloaded in the Stata format and the GPS dataset can be found when scrolling down. The name of the data set will be similar as shown in the figures below.

Zambia: Standard DHS, 2018

Select files individually or by file format type then press the button below the list of files to start the download process.

☐ ALL STATA
 ☐ ALL FLAT
 ☐ ALL SAS
 ☐ ALL SPSS
 ☐ ALL HIERARCHICAL

Survey Datasets

File Name	File Size	File Format
Births Recode		
<input checked="" type="checkbox"/> ZMBR71DT.ZIP	5.42 MB	Stata dataset (.dta)
<input type="checkbox"/> ZMBR71FL.ZIP	5.65 MB	Flat ASCII data (.dat)
<input type="checkbox"/> ZMBR71SD.ZIP	10.0 MB	SAS dataset (.sas7bdat)
<input type="checkbox"/> ZMBR71SV.ZIP	5.78 MB	SPSS dataset (.sav)
Couples' Recode		
<input type="checkbox"/> ZMCR71DT.ZIP	3.84 MB	Stata dataset (.dta)
<input type="checkbox"/> ZMCR71FL.ZIP	4.22 MB	Flat ASCII data (.dat)
<input type="checkbox"/> ZMCR71SD.ZIP	5.67 MB	SAS dataset (.sas7bdat)
<input type="checkbox"/> ZMCR71SV.ZIP	4.36 MB	SPSS dataset (.sav)

Figure 2: Download birth data (BR recode, Stata format)

After the zipped files are downloaded, they should be unzipped and slotted manually into the prespecified

Geographic Datasets		
File Name	File Size	File Format
Geographic Data		
<input checked="" type="checkbox"/> ZMGE71FL.ZIP	140 KB	Shape file (.shp)
Geospatial Covariates		
<input type="checkbox"/> ZMGC72FL.ZIP	1010 KB	Comma delimited data (.csv)

Figure 3: Download survey GPS data

folders. The BR recode data in Stata format need to go in the folder ‘Data/country/dhsStata’. A subfolder need to be created with the same name as shown on the website. In the Zambia example, it is ‘ZMBR71DT’, and unzipped files go under it. The GPS data should be placed in the ‘Data/country/dhsFlat’ folder. Similarly, unzipped files go into a subfolder with name shown on the website. In Zambia example, it is ‘ZMGE71FL’. The correct folder structure is shown in the following screenshots.







> DHS-SAR > Data > Zambia > dhsStata > ZMBR71DT				▼	↺	🔍
Name	^	Date modified	Type	Size		
 ZMBR71FL.DCT		1/29/2020 1:54 PM	DCT File	43 KB		
 ZMBR71FL.DO		1/29/2020 1:54 PM	DO File	205 KB		
 ZMBR71FL.DTA		1/29/2020 2:28 PM	DTA File	47,906 KB		
 ZMBR71FL.FRQ		1/29/2020 2:21 PM	FRQ File	3,419 KB		
 ZMBR71FL.FRW		1/29/2020 2:16 PM	FRW File	3,419 KB		
 ZMBR71FL.MAP		1/29/2020 1:54 PM	MAP File	387 KB		

Figure 4: Place survey data (BR recode) in folder

> DHS-SAR > Data > Zambia > dhsFlat > ZMGE71FL				
Name	Date modified	Type	Size	
DHS_README	12/4/2017 7:07 PM	Text Document	5 KB	
GPS_Displacement_README	2/18/2020 2:34 PM	Text Document	7 KB	
ZMGE71FL.cpg	2/19/2020 9:22 PM	CPG File	1 KB	
ZMGE71FL.dbf	2/19/2020 9:22 PM	DBF File	1,705 KB	
ZMGE71FL.prj	2/18/2020 2:34 PM	PRJ File	1 KB	
ZMGE71FL.sbn	2/19/2020 9:22 PM	SBN File	6 KB	
ZMGE71FL.sbx	2/19/2020 9:22 PM	SBX File	1 KB	
ZMGE71FL.shp	2/19/2020 9:22 PM	SHP File	15 KB	
ZMGE71FL.shp	2/20/2020 7:23 PM	XML Document	136 KB	
ZMGE71FL.shx	2/19/2020 9:22 PM	SHX File	5 KB	

Figure 5: Place survey GPS data in folder

4 GADM shapefile

The boundary shapefiles are needed for constructing subnational U5MR estimates. We use GADM shapefiles for the geography of Admin-1 and Admin-2 regions. The user could visit https://gadm.org/download_country_v3.html and enter the country of interest. By clicking ‘Shapefile’ as shown below, the zipped shapefiles will be downloaded. The preparation is completed by slotting the unzipped file in the folder ‘Data/country/shapeFiles_gadm’. An example folder structure is shown below. One should check the GADM shapefile is consistent with desired definition for Admin-1 and/or Admin-2 regions. If not, the shapefiles need to be prepared separately.

Download GADM data (version 3.6)

Country

Zambia

Geopackage

Shapefile

R (sp): level-0, level1, level2

R (sf): level-0, level1, level2

KMZ: level-0, level1, level2



Figure 6: Download GADM shapefile from GADM website

DHS-SAR > Data > Zambia > shapeFiles_gadm						
Name	Date modified	Type	Size			
gadm36_ZMB_0.cpg	9/27/2019 5:11 PM	CPG File	1 KB			
gadm36_ZMB_0.dbf	9/27/2019 5:11 PM	DBF File	1 KB			
gadm36_ZMB_0.prj	9/27/2019 5:11 PM	PRJ File	1 KB			
gadm36_ZMB_0.shp	9/27/2019 5:11 PM	SHP File	161 KB			
gadm36_ZMB_0.shx	9/27/2019 5:11 PM	SHX File	1 KB			
gadm36_ZMB_1.cpg	9/27/2019 5:11 PM	CPG File	1 KB			
gadm36_ZMB_1.dbf	9/27/2019 5:11 PM	DBF File	9 KB			
gadm36_ZMB_1.prj	9/27/2019 5:11 PM	PRJ File	1 KB			
gadm36_ZMB_1.shp	9/27/2019 5:11 PM	SHP File	314 KB			
gadm36_ZMB_1.shx	9/27/2019 5:11 PM	SHX File	1 KB			
gadm36_ZMB_2.cpg	9/27/2019 5:11 PM	CPG File	1 KB			
gadm36_ZMB_2.dbf	9/27/2019 5:11 PM	DBF File	74 KB			
gadm36_ZMB_2.prj	9/27/2019 5:11 PM	PRJ File	1 KB			
gadm36_ZMB_2.shp	9/27/2019 5:11 PM	SHP File	601 KB			
gadm36_ZMB_2.shx	9/27/2019 5:11 PM	SHX File	1 KB			
license	9/27/2019 5:11 PM	Text Document	1 KB			

Figure 7: Place GADM shapefile in folder