README

May 2011

The datasets contain baseline data from Kenya, collected in 2010, from the Measurement, Learning & Evaluation project which is the impact evaluation component of the Urban Reproductive Health Initiative.

Six datasets are available:

1. Women's dataset, filename: kenya_women_public.dta
2. Men's dataset, filename: kenya_men_public.dta
3. Household dataset, filename: kenya_hh_public.dta
4. Household roster, filename: kenya_hroster_public.dta
5. Births file, filename: kenya_borth_public.dta
6. Births file (flat), filename: kenya_birth_flat_public.dta

The women's dataset contains data from a large, representative sample of women in three intervention cities (Nairobi, Mombasa, Kisumu) plus two comparison cities (Machakos and Kakamega) in Kenya. All women ages 15-49 in selected households were eligible for interview. This dataset includes information on demographics, reproduction, contraception, maternal and child health, fertility preferences, media access and use, migration and gender relations.

The men's dataset contains data from a large, representative sample of men in the three intervention cities (Nairobi, Mombasa, Kisumu). All men ages 15-59 in selected households were eligible for interview. This dataset includes information on demographics, reproduction, contraception, fertility preferences, media access and use, and gender relations.

The household roster and births matrix are separate datasets that can be merged with household or individual level data.

The five datasets are aggregated, containing all completed interviews in all cities (only three cities for men). The table below shows the total number of completed interviews, by city for women and men. In the datasets, each city has been given a numerical value, which is also indicated in table.

Table: Number of interviews by city and city codes

City	Number of women interviewed	Number of men interviewed	City code
Nairobi	2706	1275	1
Mombasa	1465	678	2
Kisumu	1603	557	3
Machakos	1834	NA	4
Kakamega	1324	NA	5
Total	8932	2510	

There is a variable called "iestatus", which is the formal/informal distinction for that cluster. The variable named "clustid" is a masked representation of the cluster. The variable called "hhnum" is the household number within each cluster and the variable named "line" is a unique number for each member of the household. A unique identification number for individuals can be created by combining the following variables: city, clustid, hhnum, and line. Household data can be merged with individual level data by creating a unique identifier at the household level, combining city, clustid, and hhnum.

Analyses with these datasets should be weighted. To run an analysis on the aggregated data set (all five cities together), use the wweight . For example:

svyset [pweight= wweight]

To run a city specific analysis from the dataset, you would need to use the weight for that city.
Nairobi = wweight_overall_city_1
Mombasa = wweight_overall_city_2
Kisumu = wweight_overall_city_3
Kakamega = wweight_overall_city_4
Machakos = wweight_overall_city_5

For example, if you wanted to look at Nairobi data only
svyset [pweight= wweight_overall_city_1]

NOTE: In the above example, you would want to limit whatever analysis you were doing to only women in Nairobi, since wweight_overall_city_1 for everyone outside of Nairobi = 0. Ex (svy: tab age if urban==1)

Similarly, for the men's data, replace "wweight" with "mweight" and wweight_overall_city_x with mweight_overall_city_x

Key variables

Contraceptive use:

Variable name	Variable description
ever_use_fp	Any use of contraception (from Q302)
ever_use_modern_fp	Any use of modern methods
current_use	Any contraceptive use (1=Yes, 0=No)
curr_use_met	Method mix among all contraceptive users
curr_user_modern_mix	Method mix among modern contraceptive users
current_use_mix	Method mix among total population, including non-users

Knowledge of contraception: know_fp indicates whether people have knowledge of any FP method.

Sexual Activity: Women are defined as sexually active if they have had sex within the past year with the variable sexually_active.

Future Pregnancy Intention: Women's responses to Q602 and Q603 were grouped into categories to determine if they had a need to space or limit. The categories are "wants now/up to 2 yrs", "wants in more than 2 yrs", "wants after marriage", "wants/dk or missing when", "does not want anymore", "can't get pregnant", "Missing".

Unmet need: Unmet need for spacing: Includes women who are fecund and not using family planning and who say they want to wait two or more years for their next birth, or who say they are unsure whether they want another child, or who want another child but are unsure when to have the child.

In addition, unmet need for spacing includes pregnant women whose current pregnancy was mistimed, or whose last pregnancy was unwanted but who now say they want more children. Unmet need for spacing also includes amenorrhoeic women whose last birth was mistimed, or whose last birth was unwanted but who now say they want more children.

Unmet need for limiting: Includes women who are fecund and not using family planning and who say they do not want another child. In addition, unmet need for limiting includes pregnant women whose current pregnancy was unwanted but who now say they do not want more children or who are undecided whether they want another child. Unmet need for limiting also includes amenorrhoeic women whose last birth was unwanted but who now say they do not want more children or who are undecided whether they want another child.

Variable name: unmet_need_time, overall_unmet_need

Wealth: We used the same procedure as the DHS as outlined by Filmer and Pritchett, a principal component analysis (PCA) including household assets and housing characteristics. Household assets examined are categorized as a binary 'yes, have the asset' or 'no, don't have' – this doesn't include quality or comparability of a given asset that might differ across households. Further, environmental circumstances of the household (type of toilet, walls, roof and floor along with source of water) were classified as 'improved' or 'not improved' The following asset variables were included in the wealth index: vehicle, computer, TV, bicycle, dom_help (domestic help), clock, refrigerator, stove_elec (electric stove), mos_net (mosquito net), VCR, iron, sofa, torch, rooms (continuous variable on number of rooms in the house), fuel (type of fuel for cooking), separate_kitchen, floor, walls, insur (possession of medical insurance), electricity, and toilet. Wealth quintiles were created for each of the five cities individually. Additionally, a wealth quintile was created for the aggregated dataset. We have also included the component scores from the PCA (f1) for each city and overall.
Variable names (women's data): wealth_overall, wealth_nairobi, wealth_mombasa, wealth_kisumu, wealth_machakos, wealth_kakamega. Component scores: overall_score, nairobi_score, mombasa_score, kisumu_score, machakos_score, kakamega_score

Education: The education variable is a combination of Q104, Q105 and Q106. Respondents were classified as having no education, primary incomplete, primary complete and secondary plus. Adjustments for respondent's age (completion of 'form year' changed in the mid 1980s) were made.

CMC Variables: Both 'birth_order_intent' and 'mothage_grp' are variables built off of the women's CMC (century month code). Birth_order_intent is the women's intention at last birth and mothage_grp is the mother's age at last birth in categories.