Note on DHS standard weight de-normalization

By Ruilin Ren, PhD. Senior sampling statistician, ICF International

There are several sets of standard weight variables in the DHS recode: HV005 (household standard weight), V005 (women's individual standard weight) and MV005 (male's standard weight), plus HIV05 for the HIV testing weight if applicable. All the weight variables in the DHS recode are relative weights which are normalized to make the total number of unweighted cases equal to the total number of weighted cases at the national level, for each set of weight. The normalization procedure consists of multiplying the survey weight by the "estimated sampling fraction" - the un-weighted total number of completed cases divided by the weighted total number of completed cases by using the survey weight, for each set of weight and at national level. Therefore, the normalization procedure is survey specific. The normalized weight is a relative weight which is valid for estimating means, proportions and ratios, but not valid for estimating totals, and not valid for pooled data from different surveys. If one needs to estimate population totals based on DHS data or to pool data from different surveys together, the standard weight should be de-normalized. For example, to de-normalize the household standard weight HV005, one should divide the household standard weight by the household survey sampling fraction, that is, the ratio of total number of households interviewed in the survey over the total number of residential households in the country at the time of the survey. The first piece of information can be obtained from the recode data, while the second piece of information (the total number of residential households in the country at the time of the survey) may not be directly available. The second piece of information is usually obtained from population projections for a period close to the time of the survey fieldwork, based on the latest population census. The de-normalized weight is very sensitive to the second piece of information, so one should guarantee that the source of information is reliable; otherwise, it can lead to erroneous statistical conclusions. A similar procedure applies to de-normalizing the women's individual standard weights (V005) and the men's individual standard weights MV005, but one needs the total population of women age 15-49 in the country at the time of the survey for de-normalizing V005, and the total population of men age 15-49 (or 59 according to eligible age of the survey) of the country at the time of the survey for de-normalizing MV005.

HV005*=HV005x(total number of residential households in the country at the time of the survey)/(total number of households interviewed in the survey)

V005*=V005×(total females age 15-49 in the country at the time of the survey)/(number of women age 15-49 interviewed in the survey)

MV005*=MV005×(total males age 15-49 (or 15-59) in the country at the time of the survey)/(number of men age 15-49 (or 15-59) interviewed in the survey)

The population in a given age group may be more difficult to obtain. In this situation, one can use the sampling fraction of households as an approximation when de-normalizing V005 and MV005. However, since in most of the DHS surveys, male survey is conducted in a subsample of households selected for the female survey, therefore the household sampling fraction for de-normalizing MV005 should be the total number of male survey households interviewed over the total number of residential households in the country at the time of the survey.

The de-normalization procedure depends only on the weight to be de-normalized, not on the indicator to be calculated. For example, if the household standard weight (HV005) is used for calculating immunization coverage for children 12-23 months, to estimate the total number of children age 12-23 with a certain kind of immunization, it is related to de-normalizing HV005, and it is not related to the total population of children age 12-23 months in the country.

*Note: It is a tradition that the DHS recode always multiplies the standard weight variable by 1,000,000 to get integer numbers, so before using the weight variable either for estimation or for de-normalization, one needs to divide the variable by 1,000,000.