Reviewer's report

Title: Population survey sampling methods in a rural African setting: measuring mortality

Version: 1 Date: 3 January 2008

Reviewer: Shea Rutstein

Reviewer's report:

Review of

"Population survey sampling methods in a rural African setting: measuring mortality"

I find this paper very confused. At first it appears to set out a scheme for setting up a representative system of demographic surveillance sites based on covering 1% of the population. Later on, it appears the paper is about sampling for surveys within a site. While I can see situations where a sample of the site's households might be desirable, usually all household in the whole site are interviewed to gather information on demographic events. Otherwise the sampling variance would be too high for national estimates. (And then why not just to a proper national sample survey?)

It states that "more complex methods [of sampling] are often perceived as being better." However, better is not defined. For lowest sampling variance simple random samples are always better (especially when stratified). Multi-stage sample design is used to lower the costs and complexity of field work. It is unusual that the cost of a consultation of a proper sampling expert would be greater than the costs of implementing a simple random selection or using an improper design.

The paper states that selecting with probability proportional to size is "an attempt to make any individual's chance of being included in the sample similar." This statement is incorrect: PPS is used to control the overall size of the sample, not to control the probability of selection. To make a self-weighting sample, the take of households in the Ultimate Area Unit needs to fixed; otherwise, sampling weights need to be calculated and used in analysis, which is a complex operation.

The paper seems to regard the 1% sample as ideal. Depending on the indicator desired, this number may be too small or large. For example, for the two provinces covered by the Immpact census, a 1% sample would give 5,122 individuals of all ages. For under-five mortality estimates using a birth history, approximately 850 mothers would be interviewed. With about 2 children born to each mother in the last five years, there would be 1700 children. If the under-five rate is 100 per thousand births, the approximate srs confidence interval would be 86 to 114, which is probably too wide and would be even wider for subgroups of
the population and if simple random sampling were not employed.
I therefore recommend that this paper not be published in its present form.

**What next?:** Reject as not sufficiently sound