

---

Subject: Re: Stratification and sampling in Haiti

Posted by [Bridgette-DHS](#) on Wed, 05 Oct 2016 10:02:56 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Here is a response from Trevor Croft and Tom Pullum:

Your first question is why, if there are 21 strata, does the variable hv022 (sample stratum number) assume values from 1 to 163? Here is a more complete answer. DHS previously used a procedure of constructing implicit strata (163 of them, for this survey) based on pairing (or in some cases groups of 3) clusters. These implicit strata were constructed within the explicit strata (the 21 strata) and were used to calculate sampling errors. DHS stopped using this procedure some years ago, but the dataset includes the constructed implicit strata variable.

Your question about representativeness below the national level often comes up. At the stratum level, there isn't really an issue. In essence, separate samples have been drawn within each stratum. Small strata tend to be over-sampled (conversely, large strata tend to be under-sampled) in order to have enough cases to be able to make good estimates of key indicators. "Representative" has two dimensions--bias and statistical uncertainty. Stratum level estimates are unbiased and have reasonable standard errors. Yes, and you can compare strata with one another as you described (the urban and rural parts of the same region), but you should check for statistical significance. If you go below the stratum level, for example to the second administrative level, generically called districts, the estimates are still unbiased, but the standard errors go way up. It is very important to include standard errors for these lower-level estimates, just as you would for categories of a covariate at the national level. If you compare two districts within the same region, it can be difficult to get a statistically significant difference because both estimates have high standard errors.

---