

---

Subject: Re: Clustered Standard Errors

Posted by [Bridgette-DHS](#) on Mon, 08 Feb 2021 13:17:48 GMT

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

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

It's great to hear that you are using DHS data in this way. We have done some related work (<https://www.dhsprogram.com/pubs/pdf/WP142/WP142.pdf>). Unfortunately I don't have time to give a detailed response, but I can make some suggestions. First, I recommend the usual svyset for pooled surveys. Second, and most important, I would recommend working with individual-level data rather than aggregating. You construct two binary variables. The first one is S, which is 0 for the first survey and 1 for the second survey (pre-and post-intervention). The other is (say) A, which is 0 in a control area and 1 in an intervention area (area=district). The difference-in-differences approach is equivalent to assessing the significance of the interaction between A and S. If you have a binary outcome Y, then in the pooled file you do a logit regression of Y on A, S, and  $AS=A*S$ . You can include other controls, because interventions are not usually assigned at random. Then look at the sign and significance of AS. That's what we did in WP142, with the Uganda 2011 and 2016 surveys. I also applied this approach to the 2005 and 2010 surveys in Rwanda ( <https://www.ghspjournal.org/content/2/3/342/tab-supplemental> ) . If you collapse the individual-level responses and use districts as units of analysis you get into various statistical issues that can be avoided with the individual-level data.

---