Subject: SPA Missing Data - "Don't know" answers
Posted by smatthews8 on Thu, 22 Jul 2021 17:34:11 GMT
View Forum Message <> Reply to Message

I am looking at several countries' data for certain equipment to determine their presence and functionality. In looking at this data, I see that there is an option for "Don't know", which in the SPA recode data, has been aggregated with the answer that the equipment is not functioning. Is there a particular reason that the response "Don't know" is considered as a "No" rather than a missing value? The following indicators are the ones I am looking at:

• (q831c_1, q831c_2) Glucometer/Glucometer test strips Part C: Is the item in working order or expired?
• (q2331b_01/q700b_01, q2331b_02/q700b_02, q2331b_03/q700b_03) Adult/child/infant weighing scale - Part B: Functioning
• (q2331b_06/q700b_06) Thermometer Part B: Functioning
• (q2331b_08/q700b_08, q2331b_09/q700b_09) Digital/Manual BP Apparatus Part B: Functioning
• (q453) Ambulance fuel availability

The "don't know" option in the SPA recode data is aggregated with the not functioning/not present options for these. Is the assumption here that if someone does not know if the item is functioning/present, then it should be treated as a not functioning/present item?

Subject: Re: SPA Missing Data - "Don't know" answers
Posted by Bridgette-DHS on Thu, 22 Jul 2021 18:54:19 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

In general, DHS keeps "don't know" or other indeterminate responses in the denominator when calculating the percent "Yes" or "No" (as generic labels). If you threw those cases out, you would in effect be re-allocating them proportionately to the "Yes" and "No" categories. Think about a political poll, for example. Say you have 1000 people in the poll; 400 support candidate A, 300 support candidate B, and 300 are undecided. If you ignored the 300 undecided cases, you would have a majority, rather than a plurality, supporting candidate A. This would clearly be a misrepresentation of the data.

I haven't checked whether DHS always retains the indeterminate responses in the denominator, but I believe that's the general practice in all the surveys and variables.

Subject: Re: SPA Missing Data - "Don't know" answers
Posted by smatthews8 on Thu, 22 Jul 2021 19:50:40 GMT

View Forum Message <> Reply to Message

My question refers to why the "don't know" answers for these variables are being counted as a "Not functioning" or "Not present" answer for certain questions in the SPA recode datasets. So, if I want to determine the percent of hospitals in a country with the indicator "Glucometer is used, observed, and in working order", the answers that are "don't know" for whether or not the glucometer is functioning are going to be counted as a "No" in the DHS recode datasets, as they are aggregated with the answer "not functioning". This would allocate the don't know answers as a "No" response in the numerator. Thus, I am wondering what the rationale is for counting a "Don't know" answer as a "No" in the numerator for the DHS-SPA recode data. This has been the case for what I have viewed in the Afghanistan 2019, DRC 2017-18, Malawi 2013-14, Nepal 2015, Tanzania 2015, and Haiti 2019 SPA recode datasets for the indicators I mentioned above.

Subject: Re: SPA Missing Data - "Don't know" answers
Posted by Bridgette-DHS on Fri, 23 Jul 2021 11:40:19 GMT
View Forum Message <> Reply to Message

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

So far as assessing service readiness is concerned, the crucial response is "Yes". If you were constructing a null hypothesis, it would be that the facility does not have the equipment. If you have an indeterminate result, such as "don't know", the conservative strategy (consistent with minimizing Type I error in conventional hypothesis testing) would be to combine "don't know" with "No", rather than with "Yes".

I think it's a matter of whether you think of "Yes" or "No" as the outcome of interest--or, equivalently, what's the null hypothesis. In this context, the outcome of interest is "Yes". We don't want to over-estimate the prevalence of the favorable outcome, so "don't know" is combined with "No". But I can imagine an alternative approach in which the outcome of interest is "No", in which case the conservative strategy would be to combine "don't know" with "Yes", as you are suggesting.