
Subject: Missing observations in Mali BR

Posted by [MasonUWMad](#) on Thu, 07 Jan 2021 18:43:17 GMT

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

Hi, I'm using the DHS datasets for Mali from 2006, 2012, and 2018 to produce child and health and welfare measurements across time. I'm then merging the DHS data with an environmental variable to observe any significant outcome.

For Mali, the combined sample size of children in the BR recode for 2006 is 33,379; for 2012 is 33,803; and 52,140 for 2018, for a combined sample of 119,322 children. My question and concern is why are so many children missing from some of the basic health and welfare metrics? How are these missing values distributed across the sample? And is there a risk of getting a non-random sampling of Mali by using the significantly reduced sample for which there are remaining observations?

For example, the mother's age (v447a) is missing 27.19% of the sample. I use mother's age as a control variable in regressions so that immediately excludes 27.19% of the sample from all of the regressions. In tabulations for low and very low birth weight, 91.05% of the sample is missing. This is because variable m19, weight at birth, has 71.16% missing values. And of the observations there, roughly 20% are "not weighed at birth". For statistics about vitamin A vaccination, 96.61% of observations are missing for variables h33m, h33d, and h33y which are Vit A vaccination date month, day, and year respectively. For "respiratory infection in the past 2 weeks", 99.29% of observations are missing. For hemoglobin levels (hw56), 90.08% of observations are missing. Child's height/age standard deviation (hw70) and weight/height standard deviation (hw72), 79.49% and 79.23% are missing, respectively.

Please let me know if these concerns are valid and that the high number of missing values within the samples skews any of the derived health statistics or if the reduced sample size is a purposeful function of the DHS survey,

Sincerely, Mason

Subject: Re: Missing observations in Mali BR

Posted by [Bridgette-DHS](#) on Fri, 08 Jan 2021 15:00:24 GMT

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

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

The BR file includes all children in the birth histories. The KR file is limited to births in the 60 months before the survey. Both files include children who died, as well as children who survived. The woman's age is v012. You don't need v447a.

In DHS data files, a dot (in Stata) means "Not applicable". It does not mean "Missing".

Subject: Re: Missing observations in Mali BR
Posted by [MasonUWMad](#) on Mon, 11 Jan 2021 11:35:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your quick response Dr. Pullum.

I am aware of the differences between the BR and KR files, but how is that relevant to the Not Applicable values?

And thank you for correcting my misunderstanding about the dot in Stata, but if you could help me a little further - what does "Not Applicable" mean in the context of child health indicators? When a child has a dot under m19 how is their weight at birth "Not Applicable"? Either a child is weighed at birth, in which case their weight should be recorded in m19, or if they were not weighed it would be recorded in m19 as 9996 (Not weighed).

The same goes for the measurements such as height/age stdv and weight/height stdv. What does "Not applicable" mean in these contexts and why does it apply to nearly 80% of the total sample?

Lastly, is the generation of Not Applicable values randomly spread across the sample? Or when I run an analysis on the sample after the Not Applicable values are removed am I still running an analysis on a random sample of Malian households?

Thank you so much for your time and assistance Dr. Pullum.

Subject: Re: Missing observations in Mali BR
Posted by [Bridgette-DHS](#) on Mon, 11 Jan 2021 14:56:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Here are some examples of "Not Applicable" (NA or a dot in Stata). Age at death (b6 or b7) is NA if a child did not die (b5=1). The line number of the child in the household survey (b16) is NA if the child died (b5=0). The height and weight of a child in the KR file (hw2 and hw3) will be NA if the child was not in the household (not living with the mother) at the time of the survey (b16=0) or the child had died (b5=0). If a question about the child only applied to children born in the past five years, then variable will be NA for children in the BR file who were born more than 5 years ago (hw1>60). If the questions only applied to the youngest child born in the past five years, it will be NA in the KR file if bidx>1. Sometimes a set of questions will only be asked for a subsample of households, for example 1/2 or 1/3 of households. There are many such examples. Sometimes a variable only applies to the youngest surviving child who is living with the mother; otherwise it is NA. Most of these restrictions are indicated by the skips and filters in the questionnaire. Subsampling, when it occurs, is described in the first chapter of the main report.

I cannot say for sure about m19 without checking, but many of the m variables are restricted to

the youngest child (bidx=1). You could check that, for example, with "tab m19 bidx, m".
