

---

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

---