A recently published paper presents a conceptual framework for describing alternative approaches to attempting to correct for bias introduced by missing data[17]. In this framework, the analysis is divided into two stages.

**Stage 1**

At the first stage, the analyst defines a criterion that determines if an individual is “included in the analysis”. A typical criterion is that an individual is included in the analysis if the percentage of variables with complete data is greater than a threshold, and otherwise excluded.

In the context of a TB prevalence survey, we have chosen a criterion that individuals are “included in analysis” only if they participated in the survey with symptom and X-ray screening, or only in symptom screening but they gave sputum samples for TB diagnosis. The logic for this is that among eligible individuals who do not participate in the survey, information is available only on area of residence, age and sex, and there is a “large block” of missing data.

The analyst can make 3 choices about how to attempt to account for the selection bias introduced when those “included in analysis” – in our case the survey participants - are not a random subset of the eligible survey population. The first is not to attempt to correct it, denoted a “complete case” approach (CC), the second is to account for it with inverse probability weighting (IPW) applied to survey participants, and the third is to account for it by multiple imputation of the large block of missing data among individuals who did not participate (MI).

**Stage 2**

The second stage considers only those individuals “included in analysis” – in the case of TB prevalence surveys, this is survey participants. As with the first stage, there are three choices, which in this context are (i) exclusion of individuals with missing data on the outcome of pulmonary TB - a “complete case” (CC) analysis (ii) exclusion of individuals with missing outcome data, as for (i), but using inverse probability weighting to correct for the missingness so that the analysis represents all survey participants (IPW), or (iii) multiple imputation of missing outcome and other data, so as to include all survey participants in the analysis (MI). Following the overall guidance provided by Seaman et al, MI is the preferred approach to Stage 2.

Method 1 is equivalent to CC/CC (complete-case approach at both Stage 1 and Stage 2), Method 2 is MI/MI, because it relies completely on multiple missing value imputation, and Method 3 is IPW/MI, because it uses IPW for Stage 1 and MI for Stage 2.