**Notes on checking missing values for management system items**

I have created a variable for “management systems”/”mng” which is the sum of “1” or “0” answers to 13 items/questions from the audits. These are mostly from information only questions (FWG question type) in the audits (not the NCP[non-compliance point] items that we had cleaned to create the outcome variables). But there are many missing values for some of the 13 items (17% or 1105 audits had 1 or more missing mng items). I noticed that some similar questions were asked in the NCP dataset which we had used (may be due to misclassification of FWG question type as NCP in some countries for some years). This task is to check whether some of the missing values are indeed in the NCP dataset that we used or I have missed some information in the FWG dataset. If so, we could recover some of the missing values. I have uploaded the management system dataset in our GitHub repository called “ManagementSystemItems”. This is what I have cleaned but there are missing values for many items. For this purpose, you could (these are some steps that I have in mind based on my stata code experience and you are welcome to use other efficient ways):

1. Look at the “ManagementSystemItems” dataset and keep only those audits (“QuestID”) that have missing values for any management system items by selecting “mngitems”<13. These would result in a new dataset of 1091 audits with missing values for some items.
2. You can then check each of the 13 management system items by searching “potential key terms” from the codebook in the NCP dataset to see whether this/similar question was stored in the “AuditItemsRaw” dataset or NCP data from BOX. If all the key terms return no observation, search for the particular mng item can be stopped.

If some of the key terms return observations in the AuditItemsRaw dataset and the number of observations are close to 6476 (or similar numbers indicating a consistent question that is likely to be included in our outcome variable of compliance rates), the search for the particular mng item can be stopped. Note that the search of key terms may return multiple Q or QLabels that include the same key term; you may need to check whether all the return items are similar or some Qs should be excluded from this consideration.

You only need to check whether the questions in the AuditItemsRaw should be considered for the management system items when the search of key terms in the AuditItemsRaw dataset returns less than 1100 observations (which may indicate misclassification for some items). In this case, you can then see whether some of the observations’ QuestID are among the QuestIDs that have missing value for management system. If they are matched, for the same QuestID, you can then fill in the missing value for management system with the value for the “Finding” from the AuditItemsRaw dataset.

1. You can repeat step 2 within the information only questions dataset called “FWG ID and Qlabels” in BOX to check whether I have previously missed any items while using R to code the management system variables. For this check, you can first select only the QuestIDs in the FWG dataset that are matched with those QuestIDs with missing values on mng items created in step 1 (i.e. we are only looking at QuestIDs with missing values in both management dataset and the FWG dataset to save time). You can then search for key terms for each of the 13 management system items in the selected FWG dataset. If the search returns observations with questions similar to the management item, you can code the “Finding” of the FWG question and fill in the missing value in management dataset. The Finding for FWG questions are texts and you can code those with “no” as “0” and “yes” with “1” to fill in the missing values in mng data. Sometimes the text are long and you can read it and code it as “0” if the note means no such practice and code it as “1” if the note indicates the presence of such mng practice.