Methods to pool predicted probability

Suppose there is an incomplete binary variable, and we impute M times. Then, for each incomplete case, we derive a posterior distribution of the predicted probability (consisting of M predicted probabilities), or M binary imputed values (consisting of 0 and 1). Furthermore, based on the observed data, we could calculate M cut-off values of the ROC curve in each imputed dataset.

I want to distinguish two ways of generating the imputed values (0 and 1). The first approach is that the imputed values are generated by Bernoulli random variables. The imputed values (0 or 1) are the realistic values for the incomplete cases as if they were observed. We will make valid inferences in this case because we consider the uncertainty.

The second approach is that the imputed values are generated by comparing the predicted probability to the cut-off value of the ROC curve. If the predicted probability of an incomplete case is larger than the cut-off value, the imputed value will be 1. Otherwise, the imputed value will be 0. In this case, personally, I think we will derive a more accurate imputation but less valid inference.

I think there are generally two strategies to pool the predicted probability. The first one is that we aggregate the predicted probability first and then make a single decision. The second one is that we make the decision in each imputed dataset and then aggregate multiple decisions into one result.

More specific methods would be:

1. Compute the mean of cut-off values of the ROC curve in each imputed dataset E_{cutoff}; Check whether the mean of M predicted probabilities is larger than E_{cutoff}.

2. Calculate the proportion of value one and compare it with E_{cutoff} for each incomplete case. The imputed values are generated by the first approach.

3. Calculate the proportion of value one and compare it with 0.5 for each incomplete case. The imputed values are generated by the second approach.