

Diary

Lorenzo Fabbri

28/05/2023

Remaining doubts:

- Should I include the covariates when imputing the exposures?
- Should I impute the covariates by cohort?

Missing values by variables

- There is a clear pattern for some covariates and some cohorts.
- Possible solutions:
 - Drop the variables with a percentage of missing values greater than a pre-specified threshold, by cohort. Perform analyses by cohort, adjusting for different covariates, and pool the results.
 - Drop the variables with a percentage of missing values greater than a pre-specified threshold, by cohort and overall. Perform a single analysis. Then, perform sensitivity analyses by keeping those variables and dropping the cohorts with a *high* fraction of missing values.

Missing values by cases

- KAN participants seem to be the ones with the largest fraction of missing values.
- Possible solutions:
 - Drop these cases and perform the analyses. Then perform sensitivity analysis by including these cases and imputing the missing values.