Data cleaning checklist:

☐ Dataset has a codebook

☐ Variables have simple short names

☐ Each patient has a unique identifier

☐ There are no duplicate records

☐ Categorical variables have been numerically coded

☐ Dates all have the same format

☐ Missing values have been confirmed

☐ Outlier, invalid values and mixed datatypes have been identified and corrected in
continuous variables

☐ All categories have been defined for each numerically coded categorical variable

☐ If two dates are provided calculate the difference between dates to screen for errors