JPS Health Network
Office of Clinical Research

Office of Clinical Research | 1500 S. Main Street | Fort Worth, TX 76104

DATA MANAGEMENT CHECKLIST

DATA MANAGEMENT CHECKLIST:

ANALYST:
DATE:
INVESTIGATOR:
PROJECT:

☐ It is good practice to look at each relevant variable's distribution. (e.g. histograms, boxplots, percentiles, densities, summary statistics).

☐ Check for duplicates
  o Consider: If there is a duplicate ID, is the whole row of data duplicated?
  o Record number of duplicates and how they were handled.

☐ Check for nesting or repeated measures
  o Make note and use appropriate statistical design to account for it

☐ Check for missing values
  o Which variables contained missing values? For each of these variables answer:
    ▪ What percentage of values were missing?
  o What percentage of rows contained at least one missing value?
  o How were missing values handled? (e.g. removed, imputed – state method for imputation)

☐ Subset data to match investigator's exclusion/inclusion criteria

☐ Check variable classes
  o Example: a character variable that should be numeric

☐ Check for outliers
  o Consider: Checking for outliers with continuous or numeric variables might be obvious, but remember to check all variables (e.g. categorical, factor, etc.)
  o Which variables contained outliers? For each of these variables answer:
    ▪ What percentage of values were outliers?
    ▪ How were outliers handled?

☐ Check for implausible values
  o Example: a negative value for height
  o Please list any changes

☐ Check assumptions for each statistical test/model

☐ Note any additional comments or concerns.