**Instructions**

This form should be completed by both the original programmer and the QC programmer. Each checklist item must be completed by both programmers or by just one of the programmers as indicated in the “Both or Either?” column.

**Specify the grant/analysis folder:** ki1119695/PROBIT

**Name of primary programmer:** Kannan Raghavan **Date:** 12/26/2015

**Name of QC programmer:** Harsha Reddy **Date:** 12/28/2015

| **Checklist Item** | **Programmer** | | **Both or Either?** |
| --- | --- | --- | --- |
| **Primary** | **QC** |
| 1. Confirm that all variables were either mapped or identified as not needing to be mapped in the Data Decisions Log (located in the “docs” folder). | Done | Done | Both |
| 1. Univariate data checks were performed to confirm understanding of the data collection and identify implausible values and extreme outliers. | Done |  |  |
| * 1. Review summary statistics (Min, max, 5th, 25th ,50th,75th ,95th percentiles) and histograms of continuous variables | Done | Not Done | Either |
| * 1. Review frequency distributions of categorical and count data. | Not Done | Done | Either |
| 1. Report any extreme outliers or implausible values to the Data Management Lead for a decision about how to resolve. Document any conventions in the Data Decisions Log. | Done | Not Done | Either |
| 1. For longitudinal studies, review plots of data over time. | Not Done | Done | Either |
| 1. Review scatter plots of variables that are expected to be correlated with one another. | Not Done | Done | Either |
| 1. For multinational studies, repeat 2 - 4 by country. | N/A | N/A | Either |
| 1. Review summary statistics of the difference between the original and recalculated Z-scores to confirm that these do not differ by much (i.e., almost all within ±0.2). Any that differ by more than ±0.5 should be reviewed individually to understand the source of the difference. | N/A | N/A | Either |
| 1. For categorical variables that were created by remapping original values into a standard code list, review the distinct pairs of original/standard code list values to confirm that the mapping was implemented correctly. | Done | Not Done | Either |
| 1. Review row counts in datasets before and after merging to confirm that the counts are as expected. | Done | N/A | Primary only |
| 1. Review the SAS log file(s) of the data transformation program(s) looking for any errors, warnings, repeats of by-variable statements or uninitialized variable statements. In general, there should be none. | Done | N/A | Primary only |
| 1. For any more complicated, multivariate data processing tasks, review 5-10 individual subjects manually to confirm that the transformation was applied correctly. To the extent possible, be sure to include a null case (i.e., a subject for whom the relevant data is not present), a base case (i.e., a simple case) and variations of complex cases. List the variables which are reviewed here: | Done | Done | Both |
| 1. Complete a code review | N/A | Done | QC only |
| 1. After the code review, confirm that any loss of information (e.g., collapsing of categories) is documented in the Data Decisions Log. | N/A | Done | QC only |
| 1. Confirm that the SITES dataset was created with longitude & latitude populated. | Done | Not Done | Either |
| 1. Confirm that the analysis dataset was created | Not Done | Not Done | Both |
| 1. Confirm that R code was created, and executes properly, for reading the analysis dataset (xpt format) into R. Confirm that code list variables are implemented correctly as factors with the numeric variable specifying the ordering of the factor labels. (Factor labels have an implicit order even when the factors are not ordered). | N/A | Not Done | QC only |
| 1. Confirm that the data definitions file was created and is accurate. | Done | Done | Both |

**Comments**

|  |
| --- |
| Warnings about unavailable format catalogs in run\_mkdefine.log are acceptable. This happens when the corresponding folders do not exist. |