ESCAPE-NET Data transformation quality control checklist for **longitudinal** cohorts

Please fill in and send with data

|  |  |  |
| --- | --- | --- |
|  | **Cohort name:****Name Responsible Cohort Leader:****Name person who carried out transformation:****Date:** | **Initials of person who carried out quality control** |
| 1 | Identification numbers have been pseudonymized (i.e., ID\_pa and ID\_gwas) if present |  |
| 2 | Format of dataset is .csv |  |
| 3 | Each table (lc identifiers, lc basic, lc psychosocial, etc.) is stored in a separate .csv file |  |
| 4 | Each table is in long format (i.e., each subject have data in multiple rows representing baseline and each wave (ID\_visit) – see table 2) |  |
| 5 | Linking identifiers; primary key (ID\_pa) is the first variable of each table (or second variable if ID\_coh is the first variable) followed by the secondary key (ID\_visit) |  |
| 6 | Time intervals have been converted to seconds (from hh:mm:ss) |  |
| 7 | Dataset only contains ESCAPE-NET variables in ESCAPE-NET format (no local cohort variable names or coding) |  |
| 8 | Dataset does not contain empty cells. Missing values are coded as either "-3 NA; -2 not recorded; -1 unknown" |  |
| 9 | Dataset does not contain duplicate records (rows) |  |
| 10 | Summary statistics (n, n missing, mean, min, max or proportions for phenotypic data, conventional statistics for genetic data) have been calculated for all transformed data. Documentation of this quality check is sent with transformed data files |  |
| 11 | Inclusion criteria have been described and sent with transformed data files |  |

Table 1: quality control checklist - longitudinal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID\_pa | ID\_visit | ID\_visit\_y | ID\_visit\_m | ID\_visit\_base | PA\_Gender |
| 1 | 0 | 2019 | 1 | 0 | 0 |
| 1 | 1 | 2019 | 5 | 4 | 0 |
| 1 | 2 | 2019 | 11 | 10 | 0 |

Table 2: example of long format

*Please be aware that ESCAPE-NET variables (from codebook) not included in the transformed datasets will be coded as “-2: not recorded” in the joint database.*