**CSCC Statistical Computing Request**

# DATA SETS TO BE CREATED

## Describe output data set to create

### demographics

|  |  |
| --- | --- |
| **Data Set Name** | demographics |
| **Data Set Location** |  |
| **Data Set Label** |  |
| **Data Set Structure** | One record per SubjectID |
| **Key variables of data set**  **(in sort order)** | SubjectID |
| **Notes/comments (if applicable)** |  |

### physical\_assessment

|  |  |
| --- | --- |
| **Data Set Name** | physical\_assessment |
| **Data Set Location** |  |
| **Data Set Label** |  |
| **Data Set Structure** | One record per subjectID, per event |
| **Key variables of data set**  **(in sort order)** | subjectid, eventname, eventweek, l\_mot\_neuron\_upper, r\_mot\_neuron\_upper, mot\_neuron\_upper, l\_mot\_neuron\_lower, r\_mot\_neuron\_lower, mot\_neuron\_lower, l\_leg\_strength, r\_leg\_strength, leg\_strength, leg\_strength\_asym, l\_hip\_strength, r\_hip\_strength, hip\_strength, hip\_strength\_asym, back\_palp, gait, height, weight, bmi, systolic\_blood\_pressure, diastolic\_blood\_pressure, heart\_rate |
| **Notes/comments (if applicable)** |  |

### sdoh

|  |  |
| --- | --- |
| **Data Set Name** | sdoh |
| **Data Set Location** |  |
| **Data Set Label** |  |
| **Data Set Structure** | One record per SubjectID |
| **Key variables of data set**  **(in sort order)** | SubjectID |
| **Notes/comments (if applicable)** |  |

### Analysis DS data dictionaries

|  |  |
| --- | --- |
| **Data Set Name** | combined\_dictionary.xlsx |
| **Data Set Location** |  |
| **Data Set Label** |  |
| **Data Set Structure** |  |
| **Key variables of data set**  **(in sort order)** |  |
| **Notes/comments (if applicable)** | The derived dataset dictionary will be a single workbook containing one worksheet per dataset (described in C.1.a through C.1.c) based on the specifications in Tables C.2.a. through C.2.c.   * The variable order in the data dictionary should match the variable order in the tables defined in C.2. * The worksheet names should be defined based on the dataset name (e.g., demographics\_dd, phys\_assess\_dd, sdoh\_dd).   The final dictionary output should follow the [NHLBI's BioData Catalyst Data Dictionary Requirement](https://bdcatalyst.gitbook.io/biodata-catalyst-documentation/data-management/data-submission-instructions/data-dictionary-requirement).  An example is provided in [J:\BACPAC\Statistics\Special\_Projects\DataDictionaryPracticum\Supporting-Documentation\Examples\best\_example\_data\_dictionary.xlsx]. A description of the included columns are provided in the “Guide to Columns” in the example dictionary.  The columns will be described as following:   | **COLUMN\_NAME** | **DEFINITION** | | --- | --- | | VARNAME | Variable names in the dataset | | VARDESC | Variable label in the dataset | | DOCFILE | Name of the data file | | TYPE | Defined per the NHLBI BDC requirements  We want to hardcode subjectid to be string. While it includes numbers, it will always be treated as string because it includes characters.You can use the type defined in C.2 to indicate when the value = “C” then the data type in the dictionary = “string”. | | UNITS | = the text defined between the parenthesis at the end of the variable label [e.g. PAF label=Spine length (cm), so the units is “cm”] | | RESOLUTION | = number of decimal places for relevant variables | | COMMENT1 | = the details in the Definition column of the derived variable tables in section C.2. | | COMMENT2 | = the details in the Notes column of the derived variable tables in section C.2. | | VARIABLE\_SOURCE | BLANK | | SOURCE\_VARIABLE\_ID | BLANK | | VARIABLE\_MAPPING | BLANK | | UNIQUEKEY | =”X”, if VARNAME = subjectid or eventname | | COLLINTERVAL | = “Collected in [eventname value]”, if more than one eventname is associated with a variable, then include them as semi-colon separated values (e.g., Collected in Visit 1; Visit 2) | | VALUES | = the text to the left of the semi-colon in the Values and Format column of the derived variable tables in C.2 using the pipe-delimited format as used in the derived dctionary and example document. | | N | =count of non-missing values | | MISS\_N | =count of missing values | | MISS\_PERCENT | =percent of missing values among the total number of dataset records | | DISTINCT\_N | =count the number of distinct values; if TYPE=encoded | | MIN | =minimum observed value for the variable; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column (e.g., Missing=97,99 is interpreted as changing any values that are 97 or 99 to missing so they will be counted as missing and not part of the summary statisticis).  if TYPE not equal encoded and not equal string  If TYPE=date then reformat the value to ISO8601. | | MAX | =maximum observed value for the variable; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  if TYPE not equal encoded and not equal string  If TYPE=date then reformat the value to ISO8601. | | MEAN | =mean of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | STD | =standard deviation of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | PCT25 | =first quartile (i.e., 25th percentile) of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | Median | =median of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | PCT75 | =third quartile (i.e., 75th percentile) of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | |

### Form DS data dictionaries

|  |  |
| --- | --- |
| **Data Set Name** | combined\_dictionary.xlsx |
| **Data Set Location** |  |
| **Data Set Label** |  |
| **Data Set Structure** |  |
| **Key variables of data set**  **(in sort order)** |  |
| **Notes/comments (if applicable)** | The form dataset dictionary will be a single workbook containing one worksheet per dataset.   * The variable order in the data dictionary should match the variable order in the dataset. * The worksheet names should be defined based on the dataset name (e.g., ria\_dd, w12saa\_dd).   The final dictionary output should follow the [NHLBI's BioData Catalyst Data Dictionary Requirement](https://bdcatalyst.gitbook.io/biodata-catalyst-documentation/data-management/data-submission-instructions/data-dictionary-requirement).  An example is provided in [J:\BACPAC\Statistics\Special\_Projects\DataDictionaryPracticum\Supporting-Documentation\Examples\best\_example\_data\_dictionary.xlsx]. A description of the included columns are provided in the “Guide to Columns” in the example dictionary.  The columns will be described as following:   |  |  | | --- | --- | | COLUMN\_NAME | DEFINITION | | VARNAME | Variable names in the dataset | | VARDESC | Variable label in the dataset | | DOCFILE | Name of the data file | | TYPE | Defined the same as the derived variable dictionary. We want to hardcode subjectid to be string. While it includes numbers, it will always be treated as string because it includes characters.You can use the type defined in C.2 to indicate when the value = “C” then the data type in the dictionary = “string”. | | UNITS | = the text defined between the parenthesis at the end of the variable label [e.g. PAF label=Spine length (cm), so the units is “cm”] | | RESOLUTION | = number of decimal places for relevant variables | | COMMENT1 | =Prompt variable from [varlabel&rt] dataset.  Merge the source dataset with the varlabel dataset by VARNAME | | COMMENT2 | BLANK | | VARIABLE\_SOURCE | BLANK | | SOURCE\_VARIABLE\_ID | BLANK | | VARIABLE\_MAPPING | BLANK | | UNIQUEKEY | =”X”, if VARNAME = subjectid or eventname | | COLLINTERVAL | = “Collected in [eventname value]”, if more than one eventname is associated with a variable, then include them as semi-colon separated values (e.g., Collected in Visit 1; Visit 2). | | VALUES | = concatenate the “CodedValue” and “TranslatedText” using the pipe-delimited format as used in the derived dctionary and example document. | | N | =count of non-missing values | | MISS\_N | =count of missing values | | MISS\_PERCENT | =percent of missing values among the total number of dataset records | | DISTINCT\_N | =count the number of distinct values; if TYPE=encoded | | MIN | =minimum observed value for the variable; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column (e.g., Missing=97|99 is interpreted as changing any values that are 97 or 99 to missing so they will be counted as missing and not part of the summary statisticis).  if TYPE not equal encoded and not equal string  If TYPE=date then reformat the value to ISO8601. | | MAX | =maximum observed value for the variable; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  if TYPE not equal encoded and not equal string  If TYPE=date then reformat the value to ISO8601. | | MEAN | =mean of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | STD | =standard deviation of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | PCT25 | =first quartile (i.e., 25th percentile) of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | Median | =median of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | | PCT75 | =third quartile (i.e., 75th percentile) of the non-missing values; before calculating change any special missing values to actual missing using the information to the right of the semi-colon in the Values and Format column.  =BLANK, if TYPE equal encoded or string or date | |

## Detailed description of variables to be created

### Demographics Dataset

| **best\_derv\_demographics**  **Variable** | **Label** | **T** | **L** | **Values and Format** | **Definition** | **Notes** | **QC comments** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| subjectid | Participant ID | C |  |  | =[RANDOMIZED\_STAGE1=1]SUBJECTID |  |  |
| eventname | Visit Name | C |  | Visit 0 - Baseline|Visit 1|Visit 2 | "Visit 0 - Baseline" |  |  |
| eventweek | Visit Week | N |  | 0=Baseline|12=week 12|24=week 24 | =0, if eventname="Visit 0 - Baseline"  =12, if eventname="Visit 1"  =24, if eventname="Visit 2" |  |  |
| age | Age (years) | N | 2 |  | =[CCBESTDERV]AGE |  |  |
| sex\_at\_birth | Sex at Birth | N | 8 | 1=Female|2=Male|3=Intersex|99=Unknown;Missing=99 | =[CCBESTDERV]SEX\_STUDY\_MAJORITY\_VOTE |  |  |
| gender\_identity | Gender Identity | N | 8 | 1=Male|2=Female|3=Non-Binary|4=Unknown;Missing=4 | =1, if RIA2=1  =2, else If RIA2=2  =3, else If RIA2 in (3,4,5,6)  =4, else If RIA2=99  =NULL, otherwise |  |  |
| race\_reported | Race as reported | N | 8 | 1=American Indian or Alaskan Native|2=Asian|3=Black or African American|4=Native Hawaiian or Pacific Islander|5=White|88=More than one race|97=Not Reported|99=Unknown;Missing=97|99 | =[CCBESTDERV]RACE | Based on the individual race values reported in the data management system. |  |
| race\_revised | Race derived using revised categories | N | 8 | 1=White only|2=Black and Black-Multiracial|3=Asian and Asian-Multiracial (Not Black)|4=Indigenous and Indigenous-Multiracial (Not Black, Not Asian)|5=Unknown or Not reported; Missing=5 | =1, if PRE23e="true" and PRE23a="false" and PRE23b="false" and PRE23c="false" and PRE23d="false" and PRE23f="false" and PRE23g="false"  =2, else if PRE23c="true" and PRE23f="false" and PRE23g="false"  =3, else if PRE23b="true" and PRE23c="false" and PRE23f="false" and PRE23g="false"  =4, else if (PRE23a="true" or PRE23d="true") and PRE23c="false" and PRE23b="false" and PRE23f="false" and PRE23g="false"  =5, else if (PRE23f="true" or PRE23g="true")  =NULL, otherwise |  |  |
| ethnicity\_revised | Ethnicity | N | 8 | 1=Hispanic or Latino|2=Not Hispanic or Latino|3=Unknown or Not Reported; Missing=3 | =1, if [CCBESTDERV]ETHNICITY=1  =2, else if [CCBESTDERV]ETHNICITY=2  =3, else if [CCBESTDERV]ETHNICITY in (99,97)  =NULL, otherwise |  |  |
| education | Education Level | N | 8 | 1=Did not complete secondary school or some secondary school|2=High School or Secondary School Degree Complete|3=Associate’s or Technical Degree Complete|4=College or Baccalaureate Degree Complete|5=Doctoral or Postgraduate Education | =1, if RIA3 in (1,2)  =2, else if RIA3=3  =3, else if RIA3=4  =4, else if RIA3=5  =5, else if RIA3=6  =NULL, otherwise |  |  |
| employment\_status | Employment Status | N | 8 | 1=Full-time employment|2=Not employed|3=Part-time employment | =RIA4 |  |  |
| marital\_status | Marital Status | N | 8 | 1=Married|2=Never Married|3=Domestic Partner|4=Widowed|5=Divorced or Separated | =1, if RIA5=2  =2, else if RIA5=3  =3, else if RIA5=6  =4, else if RIA5=5  =5, else if RIA5 in (1,4)  =NULL, otherwise |  |  |
| household\_size | Number of people living in household | N | 8 |  | =RIA6 |  |  |
| income | Income level based on income from all sources | N | 8 | 1=Less than $10,000|2=$10,000-$24,999|3=$25,000-$34,999|4=$35,000-$49,999|5=$50,000-$74,999|6=$75,000-$99,999|7=$100,000-$149,999|8=$150,000-$199,999|9=$200,000 or more | =RIA13  =NULL, if RIA13=97 |  |  |
| poverty\_threshold | Poverty Thresholds HHS | N | 8 | 1=Low-Income|0=Not Low-Income | \*Note: “temp\_income” is the average of the range in the “income” variable;  if income=1 then temp\_income=10000;  else if income=2 then temp\_income=17499.5;  else if income=3 then temp\_income=29999.5;  else if income=4 then temp\_income=42499.5;  else if income=5 then temp\_income=62499.5;  else if income=6 then temp\_income=87499.5;  else if income=7 then temp\_income=124777.5;  else if income=8 then temp\_income=174999.5;  else if income=9 then temp\_income=200000;  else if income=missing then temp\_income=missing;  =NULL, if temp\_income=missing  =1, else if household\_size=1 and . < temp\_income < 15060  =1, else if household\_size=2 and . < temp\_income <= 20440  =1, else if household\_size=3 and . < temp\_income <= 25820  =1, else if household\_size=4 and . < temp\_income <= 31200  =1, else if household\_size=5 and . < temp\_income <= 36580  =1, else if household\_size=6 and . < temp\_income <= 41960  =1, else if household\_size=7 and . < temp\_income <= 47340  =1, else if household\_size=8 and . < temp\_income <= 52720  =0, otherwise | Participants poverty threshold as defined in ASPE HHS: https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines |  |
| birth\_date | Date of Birth |  |  |  | =PRE9 | (YYYY-MM-DD) |  |
| race\_aian | Race: American Indian or Alaska Native |  |  | true|false | =PRE23a | Raw variable value reported in the data management system. |  |
| race\_asian | Race: Asian |  |  | true|false | =PRE23b | Raw variable value reported in the data management system |  |
| race\_black | Race: Black or African American |  |  | true|false | =PRE23c | Raw variable value reported in the data management system |  |
| race\_nhpi | Race: Native Hawaiian or Pacific Islander |  |  | true|false | =PRE23d | Raw variable value reported in the data management system |  |
| race\_white | Race: White |  |  | true|false | =PRE23e | Raw variable value reported in the data management system |  |
| race\_unknown | Race: Unknown |  |  | true|false | =PRE23f | Raw variable value reported in the data management system |  |
| race\_notreported | Race: Not Reported |  |  | true|false | =PRE23g | Raw variable value reported in the data management system |  |

### Physical Assessment Dataset

*T=Type (N/C/D) for numeric, character, or date, L=length, for character variables only*

| **best\_derv\_phys\_assess Variable** | **Label** | **T** | **L** | **Values and Format** | **Definition** | **Notes** | **QC comments** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| subjectid | Participant ID | C |  |  | =[paf\_&rt]subjectid  =[w12paf\_&rt]subjectid  =[w24paf\_&rt]subjectid |  |  |
| eventname | Visit Name | C |  | Visit 0 - Baseline|Visit 1|Visit 2 | =[paf\_&rt]eventname OR  =[w12paf\_&rt]eventname OR  =[w24paf\_&rt]eventname  \*Dataset the variable is derived from depends on which time point (we need all three time points represented in dataset) |  |  |
| eventweek | Visit Week | N |  | 0=Baseline|12=week 12|24=week 24 | =0, if eventname="Visit 0 - Baseline"  =12, if eventname="Visit 1"  =24, if eventname="Visit 2" |  |  |
| l\_mot\_neuron\_upper | Left leg upper motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicables;Missing=98 | =1, if (paf18==1) OR (paf14 >= 3 OR paf16 >= 3)  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| r\_mot\_neuron\_upper | Right leg upper motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicable;Missing=98 | =1, if (paf19==1) OR (paf15 >= 3 OR paf17 >= 3)  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| mot\_neuron\_upper | Leg upper motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicable;Missing=98 | =1, l\_mot\_neuron\_upper==1 OR r\_mot\_neuron\_upper==1  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| l\_mot\_neuron\_lower | Left leg lower motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicable;Missing=98 | =1, if (paf18==2) OR (paf14 <= 1 OR paf16 <= 1)  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| r\_mot\_neuron\_lower | Right leg lower motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicable;Missing=98 | =1, if (paf19==2) OR (paf15 <= 1 OR paf17 <= 1)  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| mot\_neuron\_lower | Leg lower motor neuron | N |  | 0=Normal|1=Abnormal Behavior|98=Not Applicable;Missing=98 | =1, l\_mot\_neuron\_lower==1 OR r\_mot\_neuron\_lower==1  =0, otherwise  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| motor\_neuron | Overall motor neuron abnormality | N |  | 1=Upper Motor Neuron|2=Lower Motor Neuron|3=Mixed|4=Normal|98=Not Applicable;Missing=98 | =3, if (mot\_neuron\_lower==1 AND mot\_neuron\_upper==1)  =2, else if (mot\_neuron\_lower==1)  =1, else if (mot\_neuron\_upper==1),  =4, else if (mot\_neuron\_lower==0 AND mot\_neuron\_upper==0)  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| l\_leg\_strength | Left leg strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (paf20==3 OR paf22==0 OR paf28==0 OR paf30==0)  =2, else if (any of paf20, paf22, paf28, paf30==2)  =1, else if all of paf20, paf22, paf28, paf30==1  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| r\_leg\_strength | Right leg strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (paf21==3 OR paf23==0 OR paf29==0 OR paf31==0)  =2, else if (any of paf21, paf23, paf29, paf31==2)  =1, else if (all of paf21, paf23, paf29, paf31==1)  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| leg\_strength | Overall leg strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (r\_leg\_strength==0 OR l\_leg\_strength==0)  =2, else if (r\_leg\_strength==2 OR l\_leg\_strenth==2)  =1, else if (r\_leg\_strength==1 AND l\_leg\_strength==1)  =NULL, if r\_leg\_strength OR l\_leg\_strength are missing for baseline  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| leg\_strength\_asym | Presence of leg strength asymmetry | N |  | 1=Presence of Asymmetry|0=No Presence|98=Not Applicable;Missing=98 | =1, if r\_leg\_strength not equal l\_leg\_strength  =0, if r\_leg\_strength==l\_leg\_strength  =NULL, if r\_leg\_strenth OR l\_leg\_strength are missing for baseline  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| l\_hip\_strength | Left hip strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (paf24==0 OR paf26==0)  =2, else if (paf24==2 OR paf26==2)  =1, else if (paf24==1 OR paf26==1)  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| r\_hip\_strength | Rigth hip strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (paf25==0 OR paf27==0)  =2, else if (paf25==2 OR paf27==2)  =1, else if (paf25==1 OR paf27==1)  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| hip\_strength | Overall hip strength | N |  | 0=Absent|1=Normal|2=Weak|98=Not Applicable;Missing=98 | =0, if (r\_hip\_strength==0 OR l\_hip\_strength==0)  =2, else if (r\_hip\_strength==2 OR l\_hip\_strenth==2)  =1, else if (r\_hip\_strength==1 AND l\_hip\_strength==1)  =NULL, if r\_hip\_strength OR l\_hip\_strength are missing for baseline  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| hip\_strength\_asym | Presence of hip strength asymmetry | N |  | 1=Presence of Asymmetry|0=No Presence | =1, if r\_hip\_strength not equal l\_hip\_strength  =0, if r\_hip\_strength==l\_hip\_strength  =NULL, if r\_hip\_strength or l\_hip\_strength are missing for baseline | Only measured in Baseline. |  |
| back\_palp | Back palpation | N |  | 0=Absent|1=Mild|2=Severe|98=Not Applicable;Missing=98 | =[paf\_&rt]paf32  =98, if not baseline (eventname=Visit 0 - Baseline) | Only measured in Baseline. |  |
| Gait | Gait | N |  | 1=Normal|2=Abnormal | =[paf\_\_&rt]paf33  =[w12paf\_&rt]w12paf7  =[w24paf\_&rt]w24paf7  \*depending on which timepoint |  |  |
| Height | Height (in) | N |  |  | =[paf\_&rt]paf2  =[w12paf\_&rt]w12paf3  =[w24paf\_&rt]w24paf3  \*depending on which timepoint |  |  |
| Weight | Weight (lb) | N |  |  | =[paf\_&rt]paf1  =[w12paf\_&rt]w12paf2  =[w24paf\_&rt]w24paf2  \*depending on which timepoint |  |  |
| Bmi | BMI | N |  |  | =(weight) / (height) ^ 2 \* 703 |  |  |
| systolic\_blood\_pressure | Systolic Blood Pressure (mmHg) | N |  |  | =[paf\_&rt]paf9  =[w12paf\_&rt]w12paf4  =[w24paf\_&rt]w24paf4  \*depending on which timepoint |  |  |
| diastolic\_blood\_pressure | Diastolic Blood Pressure (mmHg) | N |  |  | =[paf\_&rt]paf10  =[w12paf\_&rt]w12paf5  =[w24paf\_&rt]w24paf5  \*depending on which timepoint |  |  |
| heart\_rate | Heart Rate (bpm) | N |  |  | =[paf\_&rt]paf11  =[w12paf\_&rt]w12paf6  =[w24paf\_&rt]w24paf6  \*depending on which timepoint |  |  |

### SDOH Dataset

*T=Type (N/C/D) for numeric, character, or date, L=length, for character variables only*

| **best\_derv\_sdoh**  **Variable** | **Label** | **T** | **L** | **Values and Format** | **Definition** | **Notes or QC comments** |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| subjectid | Participant ID | C |  |  | =[paf\_&rt]subjectid |  |  |
| eventname | Visit Name | C |  | Visit 0 - Baseline|Visit 1|Visit 2 | =[ria\_&rt]eventname |  |  |
| eventweek | Visit Week | N |  | 0=Baseline|12=week 12|24=week 24 | =0, if eventname="Visit 0 - Baseline"  =12, if eventname="Visit 1"  =24, if eventname="Visit 2" |  |  |
| lack\_transportation | SDOH Lack of Transportation | N | 8 | 1=Yes|0=No | =RIA47 | If a participant experiences a lack of transportation. |  |
| skip\_doctor\_medication | SDOH Skip Doctor Visit or Medication | N | 8 | 1=Yes|0=No | =RIA48 | If a participant experiences skipping doctor visits or medication |  |
| buy\_food | SDOH Run out of food before money to buy more | N | 8 | 0=Never True|1=Sometimes True|2=Often True | =RIA49 | If a participant worries about running out of food before having money to buy more. |  |
| food\_last | SDOH Food did not last until you had money to buy more | N | 8 | 0=Never True|1=Sometimes True|2=Often True | =RIA50 | If a participant worries about food not lasting until having money to buy more. |  |
| food\_insecurity | SDOH Food Insecurity | N | 8 | 1=Food Insecure|0=Food Secure | =1, if (RIA49 in (1,2)) OR (RIA50 in (1,2))  =0, else if (RIA49=0 and RIA50=0) |  |  |
| shut\_utilities | SDOH Utilities company shut off service | N | 8 | 1=Yes|0=No | =RIA51 | if a participant has utilities shut down. |  |
| stable\_housing | SDOH Worry for stable housing | N | 8 | 1=Yes|0=No | =RIA52 | If a participant worries about stable housing. |  |
| emotional\_support | SDOH Available emotional support | N | 8 | 1=Yes|0=No | =RIA53 | If a participant has emotional support available. |  |
| close\_friends | SDOH Number of close friends | N | 8 | 1=0|2=1-2|3=3-5|4=6-10|5=More than 10 | =RIA54 | The number of close friends a participant has. |  |
| discrim\_race\_ethnicity | Discrimination due to Race or Ethnicity | N | 8 | 1=Never|2=Rarely|3=Any | =1, if RIA55=1  =2, else if RIA55=2  =3, else if RIA55 in (3,4,5)  =NULL, otherwise | If a participant has experienced discrimination due to race or ethnicity. |  |
| discrim\_sex\_gender | Discrimination due to Sex or Gender Identity | N | 8 | 1=Never or Not Sure|2=Rarely|3=Any | =1, if RIA56 in (1,99)  =2, else if RIA56=2  =3, else if RIA56 in (3,4,5)  =NULL, otherwise | If a participant has experienced discrimination due to sex or gender identity |  |