Column Definitions of Pre-processed File

References:

1 Leroux.pdf" and "2 Organizing and analyzing the activity data in NHANES2019nihms-1525021.pdf"—along with the supplementary material "2a glz193\_suppl\_supplementary\_material.docx"— nhanes\_mort\_act\_morph\_lab\_df dataset.

The definitions are derived from the descriptions of variables in the NHANES 2003-2006 study, the Leroux et al. papers, and the associated methodology for accelerometry and mortality data. The working dataset with (3,198 rows) is a processed version of the NHANES data linked with mortality and activity data, outlined in these papers.

Below are the definitions for each column, organized by their source and context.

**Column Definitions**

**General Identifiers and Survey Data**

* **Unnamed: 0**: an index or row identifier from the original dataset or processing pipeline, not a meaningful variable (e.g., a sequential number from 0 to 3,197 in your 3,198-row dataset). Not explicitly defined in the papers but common in exported data.
* **SEQN**: Unique subject identifier assigned by NHANES to each participant, used to link data across waves and datasets (e.g., accelerometry, mortality). [Source: Page 3, "2 Organizing..."]
* **PAXCAL**: Device calibration flag, where 1 indicates the accelerometer was calibrated (all values are 1 in your data, suggesting no calibration issues). [Source: Page 3, "2 Organizing..."]
* **PAXSTAT**: Data reliability flag, where 1 indicates reliable data (all values are 1, suggesting no quality issues). [Source: Page 3, "2 Organizing..."]
* **WEEKDAY**: Day of the week (1–7) when the accelerometer data was recorded, reflecting the day the activity was measured. [Source: Page 3, "2 Organizing..."]
* **SDDSRVYR\_x**: Survey cycle indicator, where 3 = 2003-2004 and 4 = 2005-2006, representing the NHANES wave. [Source: Page 2, "1 Leroux"; Page 3, "2 Organizing..."]
* **SourceYear**: Year of the data source, likely 2003 or 2005, reflecting the NHANES wave (mean 2003.95 suggests a mix of 2003-2004 and 2005-2006). [Inferred from SDDSRVYR\_x]

**Mortality and Eligibility Data**

* **eligstat**: Eligibility status for mortality follow-up, where 1 = Eligible (all values are 1, indicating all 3,198 participants are eligible). [Source: Page 1, "1 Leroux"]
* **mortstat**: Indicator of mortality status at follow-up, where 0 = Assumed alive, 1 = Assumed deceased (15.6% deceased in your data, matching 500/3,198). [Source: Page 1, "1 Leroux"]
* **causeavl**: Indicator of whether the cause of death is available, where 0 = No cause data, 1 = Cause data available (mean 0.998 suggests nearly all deceased have cause data). [Source: Page 1, "1 Leroux"]
* **permth\_exm**: Time in months from the MEC assessment to the follow-up mortality assessment (mean 77.27, range 1–107). [Source: Page 1, "1 Leroux"]
* **permth\_int**: Time in months from the household interview to the follow-up mortality assessment (not detailed in summary but inferred as a parallel metric). [Source: Page 1, "1 Leroux"]
* **ucod\_leading**: Underlying cause of death recode from UCOD\_113 leading causes, where:
  + 001: Diseases of the heart
  + 002: Malignant neoplasms
  + 003: Chronic lower respiratory diseases
  + 004: Accidents (unintentional injuries)
  + 005: Cerebrovascular diseases
  + 006: Alzheimer's disease
  + 007: Diabetes mellitus
  + 008: Influenza and pneumonia
  + 009: Nephritis, nephrotic syndrome and nephrosis
  + 010: All other causes
  + NA: Ineligible, under age 18, assumed alive, or no cause data. [Source: Page 1, "1 Leroux"]

**Mortality Source Data**

* **mortsrce\_ndi**: Mortality source indicator from National Death Index match (binary or flag, 1 = match). [Source: Page 1, "1 Leroux"]
* **mortsrce\_cms**: Mortality source indicator from CMS information (binary or flag, 1 = match). [Source: Page 1, "1 Leroux"]
* **mortsrce\_ssa**: Mortality source indicator from SSA information (binary or flag, 1 = match). [Source: Page 1, "1 Leroux"]
* **mortsrce\_dc**: Mortality source indicator from death certificate match (binary or flag, 1 = match). [Source: Page 1, "1 Leroux"]
* **mortsrce\_dcl**: Mortality source indicator from data collection (binary or flag, 1 = match). [Source: Page 1, "1 Leroux"]

**Survey Design Variables**

* **SDMVPSU**: Primary sampling unit for the survey design, used in complex survey analysis. [Inferred from NHANES methodology]
* **SDMVSTRA**: Stratification variable for the survey design, used in weighting. [Inferred from NHANES methodology]

**Weighting Variables**

* **WTINT2YR\_x**: Interview weight for 2-year cycle, adjusted for the subset of participants (e.g., divided by 2 for combined 2003-2006). [Source: Page 6, "2 Organizing..."]
* **WTMEC2YR\_x**: Examination weight for 2-year cycle, used for accelerometry sub-study. [Source: Page 6, "2 Organizing..."]
* **RIDAGEMN\_x**: Age in months at the interview. [Inferred from NHANES age variables]
* **RIDAGEEX\_x**: Age in months at the examination. [Inferred from NHANES age variables]

**Demographic and Health Status**

* **Age**: Age in years at the time of the survey (mean 65.45, range 49–84). [Source: Page 2, "1 Leroux"; Page 7, "2 Organizing..."]
* **BMI**: Body mass index in kg/m² (mean 28.6, categorized as underweight, normal, overweight, obese). [Source: Page 2, "1 Leroux"]
* **BMI\_cat**: Categorical BMI (e.g., underweight, normal, overweight, obese). [Source: Page 2, "1 Leroux"]
* **Race**: Race/ethnicity categories (e.g., White, Mexican American, Other Hispanic, Black, Other). [Source: Page 2, "1 Leroux"]
* **Diabetes**: Self-reported diabetes status (yes/no). [Source: Page 2, "1 Leroux"]
* **CHF**: Self-reported congestive heart failure status (yes/no). [Source: Page 2, "1 Leroux"]
* **CHD**: Self-reported coronary heart disease status (yes/no). [Source: Page 2, "1 Leroux"]
* **Cancer**: Self-reported cancer status (yes/no). [Source: Page 2, "1 Leroux"]
* **Stroke**: Self-reported stroke status (yes/no). [Source: Page 2, "1 Leroux"]
* **EducationAdult**: Educational attainment (e.g., less than high school, high school, more than high school). [Source: Page 2, "1 Leroux"]
* **MobilityProblem**: Indicator of mobility difficulty (yes/no, based on walking or stair-climbing issues). [Source: Page 2, "1 Leroux"]
* **DrinkStatus**: Alcohol consumption category (e.g., nondrinker, moderate drinker, heavy drinker, missing alcohol). [Source: Page 2, "1 Leroux"]
* **DrinksPerWeek**: Number of alcoholic drinks per week (continuous). [Source: Page 2, "1 Leroux"]
* **SmokeCigs**: Smoking status (e.g., never, former, current). [Source: Page 2, "1 Leroux"]
* **yr5\_mort**: Indicator of 5-year mortality (likely redundant with mortstat, 1 = deceased). [Inferred from study focus]

**Activity and Accelerometry Data**

* **TAC**: Total activity count, a summary of overall activity volume (mean 190,617, AUC 0.783). [Source: Page 2, "1 Leroux"; Page 14, "2 Organizing..."]
* **TLAC**: Total log(1+activity count), reflecting low/light activity (mean 2,811.7 for alive, 2,278.7 for deceased). [Source: Page 2, "1 Leroux"; Page 14, "2 Organizing..."]
* **WT**: Total wear time in minutes (mean 877.1 for alive, 891.7 for deceased). [Source: Page 2, "1 Leroux"]
* **ST**: Sedentary time in minutes (mean 1,102.4 for alive, 1,184 for deceased). [Source: Page 2, "1 Leroux"]
* **MVPA**: Total minutes of moderate/vigorous physical activity (>2020 counts/min, mean 14.7 for alive, 6.5 for deceased, AUC 0.756). [Source: Page 2, "1 Leroux"; Page 14, "2 Organizing..."]
* **SATP**: Sedentary to active transition probability (mean 0.05, AUC 0.658). [Source: Page 2, "1 Leroux"; Supplementary Material]
* **ASTP**: Active to sedentary transition probability (mean 0.29 for alive, 0.37 for deceased, AUC 0.733). [Source: Page 2, "1 Leroux"; Supplementary Material]
* **mi1**: Mean score of the first principal component from FPCA, associated with overall activity (OR 1.014). [Source: Supplementary Material]
* **si1**: Standard deviation of the first principal component score, associated with variability (not retained in final model). [Source: Supplementary Material]
* **mi2**: Mean score of the second principal component (not detailed, likely activity pattern). [Inferred from FPCA]
* **si2**: Standard deviation of the second principal component score (not retained). [Inferred from FPCA]
* **mi3**: Mean score of the third principal component (not detailed). [Inferred from FPCA]
* **si3**: Standard deviation of the third principal component score (not retained). [Inferred from FPCA]
* **mi4**: Mean score of the fourth principal component (not detailed). [Inferred from FPCA]
* **si4**: Standard deviation of the fourth principal component score (not retained). [Inferred from FPCA]
* **mi5**: Mean score of the fifth principal component (not detailed). [Inferred from FPCA]
* **si5**: Standard deviation of the fifth principal component score, associated with variability (not retained in final model). [Source: Supplementary Material]
* **mi6**: Mean score of the sixth principal component (not detailed). [Inferred from FPCA]
* **si6**: Standard deviation of the sixth principal component score, associated with lower mortality risk (OR 0.926). [Source: Supplementary Material]
* **sPC1**: Score of the first principal component (likely mean or derived metric). [Inferred from FPCA]
* **sPC5**: Score of the fifth principal component (likely mean or derived metric). [Inferred from FPCA]
* **sPC6**: Score of the sixth principal component (likely mean or derived metric, surrogate for SD). [Source: Supplementary Material]

**Morphometric and Lab Data**

* **survey\_year**: Year of the survey (2003 or 2005). [Inferred from SDDSRVYR\_x]
* **Gender**: Gender of the participant (Male/Female, 51% Male, 49% Female). [Source: Page 2, "1 Leroux"]
* **BMXWT**: Weight in kg from body measurement exam. [Inferred from NHANES BMX data]
* **BMXHT**: Height in cm from body measurement exam. [Inferred from NHANES BMX data]
* **BMXWAIST**: Waist circumference in cm. [Inferred from NHANES BMX data]
* **BMXBMI**: BMI calculated from BMXWT and BMXHT (mean 28.6). [Inferred from NHANES BMX data]
* **WHtR\_calculated**: Waist-to-height ratio, calculated as BMXWAIST / BMXHT. [Inferred from NHANES data]
* **CRP (mg/dL)**: C-reactive protein level (mean 0.33, marker of inflammation). [Inferred from NHANES lab data]
* **LogCRP (mg/dL)**: Log-transformed CRP for normality (mean -1.11). [Inferred from NHANES lab data]
* **OSA\_Probability**: Probability of obstructive sleep apnea (derived metric). [Inferred from NHANES data]
* **HTN\_SelfReport**: Self-reported hypertension status (yes/no). [Inferred from NHANES questionnaire]
* **AvgSysBP**: Average systolic blood pressure in mmHg. [Inferred from NHANES exam data]
* **AvgDiaBP**: Average diastolic blood pressure in mmHg. [Inferred from NHANES exam data]
* **HTN\_Category\_Measured**: Hypertension category based on measured BP (e.g., normal, elevated). [Inferred from NHANES data]
* **Age\_Group**: Age category (e.g., 50-59, 60-69, 70+). [Inferred from NHANES data]
* **Albumin (g/dL)**: Serum albumin level (mean 4.2). [Inferred from NHANES lab data]
* **Blood urea nitrogen (mg/dL)**: BUN level (mean 15.6). [Inferred from NHANES lab data]
* **Total calcium (mg/dL)**: Serum calcium level (mean 9.5). [Inferred from NHANES lab data]
* **Cholesterol (mg/dL)**: Total cholesterol level (mean 200.3). [Inferred from NHANES lab data]
* **Bicarbonate (mmol/L)**: Serum bicarbonate level (mean 25.1). [Inferred from NHANES lab data]
* **Glucose, serum (mg/dL)**: Serum glucose level (mean 102.4). [Inferred from NHANES lab data]
* **Lactate dehydrogenase LDH (U/L)**: LDH enzyme level (mean 120.5). [Inferred from NHANES lab data]
* **Phosphorus (mg/dL)**: Serum phosphorus level (mean 3.7). [Inferred from NHANES lab data]
* **Total protein (g/dL)**: Total serum protein level (mean 7.2). [Inferred from NHANES lab data]
* **Triglycerides (mg/dL)**: Triglyceride level (mean 150.2). [Inferred from NHANES lab data]
* **Uric acid (mg/dL)**: Uric acid level (mean 5.8). [Inferred from NHANES lab data]
* **Creatinine (mg/dL)**: Serum creatinine level (mean 0.97). [Inferred from NHANES lab data]
* **Osmolality (mmol/Kg)**: Serum osmolality (mean 279.4). [Inferred from NHANES lab data]
* **Globulin (g/dL)**: Serum globulin level (mean 2.97). [Inferred from NHANES lab data]