

DAT_CLEAN Specifications

1 Specifications for DAT_CLEAN

- Rename the first column to Ptid
- Define age.geq.65, a binary variable indicating Age ≥ 65
- Define TwophasesampInd, a binary two-phase sampling indicator for D57 analyses as the intersection of
 - getting two vaccine shots
 - either in the subchort or be a case ($\text{EventIndPrimaryD29} == 1$)
 - no missing values in baseline, Day 29, or Day 57 binding to Spike or RBD
- Define TwophasesampInd.2, a binary two-phase sampling indicator for D29 analyses as the intersection of
 - getting two vaccine shots
 - either in the subchort or be a case ($\text{EventIndPrimaryD29} == 1$)
 - no missing values in baseline or Day 29 binding to Spike or RBD
- Define ethnicity, a factor variable that has three levels: "Hispanic or Latino", "Not Hispanic or Latino", "Not reported and unknown"
- Define race, a factor variable that has the following levels: "White", "Black or African American", "Asian", "American Indian or Alaska Native", "Native Hawaiian or Other Pacific Islander", "Multiracial", "Other", "Not reported and unknown"
- Define WhiteNonHispanic, a binary variable that takes value 1 if race is White and ethnicity is not Hispanic, 0 if race is not "white or unknown" OR ethnicity is Hispanic, and NA otherwise
- Define URM, a binary variable that takes value 1 if WhiteNonHispanic is 0, and 0 otherwise including when WhiteNonHispanic is NA
- Define Bstratum, an integer variable to index randomization strata. For Moderna, there are three strata: age ≥ 65 , age < 65 and HighRiskInd==1, age < 65 and HighRiskInd==0
- Define demo.stratum, an integer variable to index the demographics strata. For Moderna, there are six strata: URM [1, 0] x Bstratum
- Define tps.stratum, an integer variable to index correlates sampling strata. For Moderna, there are 24 strata: [placebo/baseline negative, placebo/baseline positive, vaccine/baseline negative, vaccine/baseline positive] x demo.stratum
- Define Wstratum, an integer variable to index strata for computing weights. For Moderna, there are 28 strata: if not case ($\text{EventIndPrimaryD29}==1$): demo.stratum, else: [25 - placebo/baseline negative, 26 - placebo/baseline positive, 27 - vaccine/baseline negative, 28 - vaccine/baseline positive]
- Define wt, a double variable for D57 correlates analyses. The weights equal the sampling probabilities of TwophasesampInd==1 within each Wstratum. This variable is only defined for subjects satisfying Perprotocol == 1 and EventTimePrimaryD57 ≥ 7 and set to NA otherwise

- Define wt.2, a double variable for D29 correlates analyses. The weights equal the sampling probabilities of TwophasesampInd.2==1 within each Wstratum. This variable is only defined for subjects satisfying Perprotocol == 1 and EventTimePrimaryD57>=7 | EventTimePrimaryD29 >= 7 and EventTimePrimaryD29 <= 13 and Fullvaccine == 1 and set to NA otherwise
- Define wt.subcohort, a double variable for immunogenicity analyses that use subcohort only and are not enriched by cases outside subcohort. The weights equal the sampling probabilities of TwophasesampInd==1 and SubcohortInd==1 within each tps.stratum. This variable is only defined for subjects satisfying Perprotocol == 1 and EventTimePrimaryD57>=7 and set to NA otherwise
- Carry out single imputation using the R package mice for the subjects satisfying TwophasesampInd==1. Since baseline, D29 and D57 binding to Spike and RBD are required in this subset of people, the imputation is intended to fill in missing marker values for neutralization assays at these three time points. (_common.R defines a vector variable, markers, for this purpose.)
- Define Delta57overBxxx, Delta29overBxxx, and Delta57over29xxx for each xxx in the variable assays. They are the log10 fold rise markers.
- Censor each of the markers in all combinations of assays and markers. Censoring means setting values < lower limits of detection (llods) to half of the llods. Note that llods are on the antilog scale and need to be transformed.