CoR_COXPH Specifications

1 Specifications for CoR_COXPH

The population of interest for the D57 CoR study is defined as baseline negative, per-protocol, at risk at 7 days post the Day 57 visit, vaccine recipients. In the rest of this specification we will refer to this population as baseline negative vaccine recipients for simplicity. For comparison, we are also interested in the baseline negative, per-protocol, at risk at 7 days post the Day 57 visit, placebo recipients.

The population of interest for the D29 CoR study is defined similarly, but per-protocol AND at risk at 7 days post the Day 57 visit needs to replaced by (at risk at 14 days post the Day 29 visit AND per-protocol) OR (at risk between 7 and 13 days post the Day 29 visit AND per-protocol AND receiving both vaccine doses).

There are two main sets of results in the CoR_COXPH section of the report: (i) Cox regression model coefficients, and (ii) marginalized risk functions based on Cox models.

1.1 Cox regression model fits

Due to the two-phase sampling design, the survey package svycoxph function (using twophase function to create the design object) will be used for fitting Cox models and getting confidence intervals and P values.

The following continuous biomarkers will be studied. Their trichotomized version adds cat at the end, e.g. Day57bindSpikecat.

- Day57bindSpike
- Day57bindRBD
- Day57pseudoneutid50
- Day57pseudoneutid80.

We first fit Cox regression models to the baseline negative, vaccinee population (defined in more details above), one for each continuous or trichotomized marker. All models adjust for the following baseline covariates. These results will be double programmed.

- at risk or not
- community of color or not
- age (to be replace by SL risk score)

Notes:

- The No. at-risk in the trichotomized table are weighted estimates based on the phase 2 samples.
- The q-values and FWER are computed using Sue Li's implementation of the Westfall and Young permutation-based method (kyotil::p.adj.perm) for the continuous and trichotomized markers together. We verify Sue's implementation through a Monte Carlo study that is outside of this verification plan.

Second, we fit Cox regression models to the following subgroups for continuous markers only. Note that for some analyses we cannot adjust for all baseline demographic variables, e.g. in the second bullet we cannot include the at-risk indicator, HighRiskInd, as a covariate. The clinical covariates to be excluded These results will be code-reviewed.

- age >= 65
- age<65 and at risk (-HighRiskInd)
- age<65 and not at risk (-HighRiskInd)
- at risk (-HighRiskInd)
- Not at risk (-HighRiskInd)
- Comm of color (-MinorityInd)
- White non-Hispanic (-MinorityInd)
- Men
- Women

1.2 Marginalized risk

The primary implementation of marginalized risk uses functions from the package marginalizedRisk on CRAN. The marginalizedRisk package is written for a more general purpose. The verification will double program the specific algorithm used to compute the marginalized risk in this report. Because of the specialized nature of the R functions required to implement these methods, the tester will first review the primary programmer's code to gain familiarity with the basic functions for working with survival analysis in R, but will write the program independently after that.

There are four types of curves based on marginalized risks. The first three are for continuous markers and the risk here refers to the marginalized cumulative risk by (inclusive) t_F, which is defined as the time of the last observed outcome in the vaccine arm.

Marginalized risk function conditional on $S = s$ for a continuous marker:

- Fit a two-phase Cox model to the baseline negative vaccine recipients with baseline covariates and the continuous marker
- For each s in 5%, 6%, 95% percentile of the marker values (defined ignoring weights),
  - Compute the risk for all phase 2 subjects, but with their marker value set to s instead of their original S
  - Return the weighted mean of the risk

Controlled VE function conditional on $S = s$ for a continuous marker:

- Fit a two-phase Cox model to the baseline negative vaccine recipients with baseline covariates and the continuous marker
- For each s in 5%, 6%, ..., 95% percentile of the marker values (defined ignoring weights),
  - Compute the risk for all phase 2 subjects, but with their marker value set to s instead of their original S
  - Return 1- (weighted mean of the risk) / (placebo overall risk)

Marginalized risk function conditional on $S >= s$ for a continuous marker:

- For each s in 0%, 5%, ..., 90% percentile of the marker values (defined ignoring weights),
  - Fit a two-phase Cox model to the baseline negative vaccine recipients with baseline covariates using the subset of samples satisfying $S >= s$
  - Compute the risk for all phase 2 subjects, not just the subset with $S >= s$
  - Return the weighted mean of the risk

The last curve is for discrete markers, and it plots the marginalized cumulative incidence rate as a function of time for each subpopulation defined by the discrete marker. The risk here refers to the marginalized cumulative risk at a series of time points.

Marginalized cumulative incidence rate function conditional on $S = s$ for a discrete marker:

- Fit a two-phase Cox model to the baseline negative vaccine recipients with baseline covariates and the discrete marker
- For each s in 5%, 6%, ..., 95% percentile of the marker values (defined ignoring weights),
  - For each time t of interest,
    - Compute the risk by (inclusive) time t for all phase 2 subjects, but with their marker value set to s instead of their original S
    - Return the weighted mean of the risk