

Supplement 3: Statistical analysis plan

Sample size and effect size estimation

BeLOVE is an observational study that aims to address different research questions. The primary research aim is to identify risk factors for MACE using a comprehensive set of biomarkers and variables from various time points and to develop a prediction model for MACE.

It is assumed that within the 1-year follow-up 8%–12% of the patients experience a new CVE of any type. For a two-sided significance level of 5%, a power of 80%, and no correlation between the predictors and an annual event proportion of 8%, the minimal detectable effect in 7000 participants is HR 1.13, which is lowered to HR 1.10 when a 12% annual incidence of outcome is assumed. Since we assume a drop-out rate of around 12%, 8000 patients should be included.

Statistical analyses

Before analysing the primary and secondary outcomes, detailed statistical analyses plans will be provided. In general, for independent and for outcome variables, depending on their scaling, descriptive measures, such as means, SD, medians, IQRs or absolute and relative frequencies will be reported. Kaplan-Meier survival curves and estimates will be used as descriptive measures for time-to-event data. Analyses will be stratified by index disease entity.

Cox proportional hazards regression, adjusted for potential confounders, will be used for the primary analysis to model the association between the exposure variables of interest and the major endpoint (MACE). Time from inclusion into the study or acute event will be used to calculate person-time at risk until the occurrence of MACE, competing events or censoring. Among others, we will use directed acyclic graphs to model the path from exposure to event, and to identify potential confounding or intermediary variables.

Additionally, statistical regression models for recurrent or competing events and structural equation models for time-to-event data will be used to analyse more thoroughly the relations of different markers and characteristics to the outcomes. Continuous outcomes (e.g., plasma levels of biomarkers), binary endpoints or ordinal endpoints will be analysed using linear, generalised linear, binary logistic or ordinal regression models. Analyses will be stratified by disease entity of the index event/disease.

Effect estimates and corresponding CIs will be reported where possible and appropriate, instead of reporting 'statistical significance', as is recommended by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline for the reporting of observational studies.