Please note: This is an updated statistical analysis plan with minor modifications.

STATISTICS AND METHODOLOGY

Statistical analysis plan for nested cohort study:
Mixed effects models with optimal type of mixed-effects model (e.g. mixed-effects linear model or mixed-effects beta-regression) determined by model fit. Exposed patients (infected with organism of interest) and unexposed patients will be matched on hospital, and time in hospital before index date.

The analysis will include fixed effects for the matching variables and the following additional covariates: age, sex, comorbidities (Charlson Comorbidity Index), surgical procedure within 30 days before the index date (date of matching), antibiotic use within 30 days before the index date. Time will also be included as a covariate to model changes over time, with an interaction with the exposures of interest to model potential time-varying effects of the exposure. Total quality of life losses will be estimated and compared by obtaining the area under the curves for exposed and unexposed groups using Simpson's rule (quadratic interpolation).

A cluster-specific and patient-specific random effect will be considered to model the repeated measurements on the same cluster and patient. Supportive analyses considering more complex random effects structures will also be investigated. (e.g., time within clusters, wards within hospitals). The interaction between time and interventions will also be added as a fixed effect to model a possible time-varying intervention effect.

It is possible that a limited number of individuals that are recruited as uninfected controls will attract a CRE/CRPA/CRAB infection at a later point during their hospitalisation. This is necessary to avoid bias introduced when selecting controls that will never be infected (conditioning on the future). In expectation, the number of people acquiring such infections is small and measurements on or after the day of infection in those patients originally assigned to the control group will be censored.

Sample size calculation for nested cohort study
The RCT this matched cohort study will be nested in, will be conducted in 24 acute care hospitals with ~75 infections per year caused by carbapenem-resistant enterobacteriales (CRE), carbapenem-resistant Pseudomonas aeruginosa
(CRPA), and carbapenem-resistant Acinetobacter baumannii (CRAB). The study will have a duration of 4 years, leading to approximately 24*75*4 = 7,200 CRE/CRPA/CRAB infections. Assuming a similar distribution of types of infections – bloodstream, urinary tract infection, etc. – as observed in point prevalence surveys among hospitals across Europe, a drop-out of 10% over time, and at least 80% power to detect a difference in utility of 0.05 at each time-point for all HAI of interest and a difference of 0.1 for bloodstream infections one would need to recruit approximately 4,500 participants (189 – 63 infected and 126 uninfected – patients per hospital over 4 years).

Handling of missing data and drop-outs
Missing data
Quarterly checks on data completeness with feedback to the centres will be organised. Delays or errors of data collection will be discussed in the quarterly videoconferences with the hospitals.

Drop-outs

Hospitals dropping out of the study will be replaced until month 9, which allows a minimum baseline of 6 months. Thereafter, hospitals dropping out will not be replaced. The power calculation is conservative, and the primary outcome can still be analysed.