Effectiveness of a **P**ressure **I**njury Clinical Judgment **RI**sk-screening tool and prevention plan: A **C**luster **R**andomized **C**ontrol **T**rial (EmPiRIC RCT)

Statistical Analysis Plan

cRCT - The sample will be summarized descriptively. Exploratory analyses will be conducted on the data to check for data distributions, patterns and extent of any missing data, outliers and influential points. Data transformations will be considered if data is skewed or shows other gross violations from normality. The missing data will be assessed by methods such as separate variance t-tests and Little's chi-squared tests, and appropriate strategies utilized accordingly (e.g. complete case analysis, deletion of variables, imputation by expectation maximization etc.)

The analysis of the primary outcome will be conducted using hierarchical regression models, in which patients represent the lower level and wards the upper level of the analysis. Random intercept models will be conducted with appropriate adjustments made to standard errors and confidence intervals due to clustering. Data will be transformed using the logistic transformation. The intra-class correlation will be evaluated using a simulation method to evaluate level-1 variance. Group balance across key covariates will be assessed and any covariate for which a substantive imbalance exists will be included in the model as a control variable alongside the key treatment variable, which will be forced into all models. P-values, odds ratios and associated 95% confidence intervals will be reported for all parameters.

The secondary analysis of time to HAPI occurrence within a 30-day period will be analysed using fully parametric time-to-event methods using interval censoring where the precise time of occurrence of a HAPI is not known. The model will use the same 2-level hierarchy, with the same corrections for data clustering, as described for the analysis of the primary outcome. Any variable included in the final multilevel logistic regression model above will also be fitted into the time-to-event model. Several candidate modelling distributions will be assessed for goodness-of-fit to data (e.g. exponential, Weibull, log-logistic etc.) with the best fitting distribution chosen by consideration of AIC statistics. P-values, hazard ratios or acceleration factors, with associated 95% confidence intervals, will be reported for all parameters and survival curves of individuals discriminated by variables of substantive importance will also be presented.

The secondary analysis of length of patient stay will be analysed using regression modelling under the same 2-level hierarchy, with the same corrections for data clustering, as described for the analysis of the primary outcome. P-values and parameter estimates with associated 95% confidence intervals, will be reported for all parameters.

Supplementary ward-level analyses will also be undertaken, to determine cumulative incidence and/or incidence rates in each type of ward in both groups.

Data will be analysed using Stata statistical software and MLwiN multilevel modelling software.

Focus groups: Thematic analysis (Braun & Clarke, 2014) will be used to identify, analyse and report themes within the focus group data. All qualitative data will be organised and analysed using NVivo software.