SUPPLEMENTARY MATERIALS

Statistical Methods

1 Pre-trial sample size calculation – further details

To allow for the longitudinal nature of the trial the following additional assumptions were made:
Each ICU ward provides 14 beds. Over the initially planned duration of recruitment of 12 months, an estimated average ICU length of stay of 3 days would result in a total of **1703 *cases*** per year in one ICU (approx. (365:3)·14). One *case* is defined as a patient's ICU stay.
There is little information to support likely values for the proportion of eligible patients with statutory health insurance coverage who can be enrolled into a quality improvement intervention study: It is anticipated that 204 cases per ICU per year can be recruited. Based on expert knowledge, about 20% of cases are re-admissions resulting in a cluster size of about 163 patients (~204·0.8). Therefore, 9 clusters with a total of 1467 cases have to be allocated to the trial (3 clusters per sequence group).
In order to compensate for withdrawal of clusters, cluster attrition or poor participant accrual (conservative enrolment of patients admitted to the ICU by site personnel might reduce recruitment rates), one additional cluster per sequence group will be added. Thus, assuming a balanced, equal number of clusters in each sequence, 12 clusters have to be randomized aiming to enrol a total of **1956 cases** (12·163).

2 Clinical evaluation – further details

Principal model for primary outcomes

The primary analyses for the clinical evaluation aim to determine if there is a difference between control and intervention proportion with respect to the adherence to a single QI. A primary efficacy outcome (single QI fulfilled yes/no at a specific date for a certain patient) is measured on the patient-level continuously in time from ICU admission until discharge from ICU.

All eight co-primary outcomes will adjust for calendar time since the intervention condition is sequentially rolled out and patients within the same randomization unit are not independent. To be more detailed, an underlying *secular trend (calendar time),* i.e. time since the start of the study, will be assumed which may be caused by temporal trends or periodicity/seasonality unrelated to the study; additionally, the length of *exposure time* varying by cluster needs to be accounted for in the analyses considering a cumulative effect on outcomes.[1-3]

Specification of the principal model to estimate the degree of performance with respect to a single QI

Estimation of the intervention effect will be obtained from a logistic regression model with log link which comprises random intercepts for cluster and patient, and tele-ICU intervention (i.e. a binary variable for ERIC intervention vs. control condition) as fixed effect. Besides, confounding of the intervention effect with time has to be considered; hence, time has to be adjusted for irrespective of statistical significance.[4] The model will separately adjust for **calendar time** (time since start of recruitment within a single cluster) and **exposure time** (time since roll-out of the intervention; 0 otherwise) as well as exposure to the **training phase** (time since start of the training activities; 0 otherwise).

For each QI, the same generalized linear mixed effects model (GLMM) will be fitted within a continuous-time framework. The relative effect sizes (odds ratios for QI adherence on ERIC intervention vs. usual care) with associated 95% confidence intervals (CIs) will be reported to quantify

intervention effects, together with variance components to inform planning of future trials. Multiple imputation techniques will not be considered while assuming missingness at random (MAR).[5]

In supplementary analyses for the primary efficacy outcomes a nonparametric model using spline functions [6] for continuously measured time variables will be conducted in order to explore nonlinear trends, time-varying effects of the intervention, lag effects, or time-by-treatment interactions (e.g., the effect of the intervention might be quite large right after the roll-out with its impact starting to decline over increasing time since exposure).

Other additional analyses

Further, an assessment of whether the intervention effect varies between sequence groups (defining the date for site transitioning) will be investigated within a supplementary analysis by extending the principal model by an additional fixed effects term.

Analyses of secondary outcomes

Analyses for secondary outcomes concerning ICU or post-ICU follow-up (measurements assessed at months 3 and 6) will be conducted using GLMM for categorical outcomes and the logit link for binary outcomes, reported as odds ratios. In the case of continuous secondary outcomes an identity link (resulting in a linear regression model) will be applied, and the difference in means between both conditions will be reported.

For key secondary outcomes (e.g., PICS-related patient-centered functional outcomes derived from patient questionnaires or tests) assessed during both follow-up visits, a high proportion of missing data is anticipated due to the high severity of illness in this study population.[7] Therefore, functional outcomes might be truncated due to death for ICU patients exposed to the control condition if the intervention improves survival. Thus, worse or unchanged functional outcomes on intervention condition might be the result of survival of patients who, without the intervention, would have died.

Statistical methods suitable for the analysis of randomized controlled trials when the intervention may reduce mortality, but where functional outcomes are also measured, will be applied as proposed by Colantuoni and colleagues.[8] Depending on the fraction of missingness, methods constructing a composite endpoint that include both mortality and functional outcomes will be applied as *sensitivity analyses* to compare between both conditions and to derive estimates for the effect of the intervention 3 or 6 months after index ICU discharge (for details refer to Lachin [9]).

References

1. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M *et al*: **Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration**. *BMJ* 2018, **363**:k1614.

2. Nickless A, Voysey M, Geddes J, Yu L-M, Fanshawe TR: **Mixed effects approach to the analysis of the stepped wedge cluster randomised trial—Investigating the confounding effect of time through simulation**. *PLoS One* 2018, **13**(12):e0208876.

3. Barker D, McElduff P, D'Este C, Campbell MJ: **Stepped wedge cluster randomised trials: a review of the statistical methodology used and available**. *BMC Med Res Methodol* 2016, **16**(1):69.

4. Hussey MA, Hughes JP: **Design and analysis of stepped wedge cluster randomized trials**. *Contemp Clin Trials* 2007, **28**(2):182-191.

5. O'Kelly M, Ratitch B: **Clinical Trials with Missing Data: A Guide for Practitioners**: John Wiley & Sons; 2014.

6. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M: **A review of spline function procedures in R**. *BMC Med Res Methodol* 2019, **19**(1):46.

7. Wilcox ME, Ely EW: **Challenges in conducting long-term outcomes studies in critical care**. *Curr Opin Crit Care* 2019, **25**(5):473-488.

8. Colantuoni E, Scharfstein DO, Wang C, Hashem MD, Leroux A, Needham DM, Girard TD: **Statistical methods to compare functional outcomes in randomized controlled trials with high mortality**. *BMJ* 2018, **360**:j5748.

9. Lachin JM: **Worst-rank score analysis with informatively missing observations in clinical trials**. *Control Clin Trials* 1999, **20**(5):408-422.