SAP for Final Analysis

SETPOINT 2

Stroke-related Early Tracheostomy vs. Prolonged Orotracheal Intubation in Neurocritical care Trial 2

ClinicalTrials.gov NCT02377167

Signatures

Responsible Biometrician
Prof. Dr. Meinhard Kieser
meinhard.kieser@imbi.uni-heidelberg.de

08.12.2020
Date / Signature

Representatives of the responsible Biometrician
Dr. Laura Benner
benner@imbi.uni-heidelberg.de

08.12.2020
Date / Signature

Jan Meis
meis@imbi.uni-heidelberg.de

08.12.2020
Date / Signature

Principal Investigator
Prof. Dr. Julian Bösel
Julian.Boesel@klinikum-kassel.de

08.12.2020
Date / Signature

Co-Principle Investigator
Dr. David Seder
sederd@mmc.org

08/12/2020
Date / Signature

Medical Coordinator
Dr. Silvia Schönenberger
Silvia.Schoenenberger@med.uni-heidelberg.de

9.12.2020
Date / Signature

Contents

Abbreviations

Abbreviation	Definition
AE	Adverse event
AIS	Acute ischemic stroke
eCRF	Electronic case report form
FCS	Fully conditional specification
GCS	Glasgow Coma Scale
ICH	Intracerebral hemorrhage
ICP	Intracranial pressure
ICU	Intensive care unit
IMBI	Institute of Medical Biometry and Informatics
mRS	Modified Rankin Scale
NCCU	Neurocritical care unit
NIHSS	National Institute of Health Stroke Scale
PP	Per protocol
SAE	Serious adverse event
SAH	Subarachnoid hemorrhage
TT	Tracheostomy

1 Purpose of the SAP

In the trial protocol of the present study, several assumptions, methods and procedures of the statistical analysis of the SETPOINT2 study have been described. The statistical analysis plan (SAP) aims to specify in detail the different analysis sets and its criteria, the calculation of outcome measures based on observed data and evaluations carried out in the statistical analysis after study completion of the SETPOINT2 study. An interim analysis was conducted after the first 127 enrolled patients (33% of the total planed enrolment). The null hypothesis for the primary endpoint could not be rejected at the pre-specified significance level and therefore, the study was not stopped for efficacy. This SAP describes the analysis strategies of the final analysis, where the planned total sample size was reached.

2 Objective of the Trial

Tracheostomy is a common procedure in long-term ventilated critical care patients and frequently necessary in those with severe stroke. The optimal timing for tracheostomy is still unknown, and it is controversial whether early tracheostomy impacts upon functional outcome. Patients are randomized to either percutaneous tracheostomy within the first five days after intubation (“early tracheostomy”) or to ongoing orotracheal intubation with consecutive weaning and extubation and, if the latter failed, to percutaneous tracheostomy from day 10 after intubation (“prolonged intubation”). The primary aim of the trial is a comparison between both groups with respect to the functional outcome defined by the modified Rankin Scale (mRS dichotomized to 0-4 vs. 5-6), and in a hierarchical manner the overall death (mRS = 6) six months after admission to the ICU will be compared. SETPOINT2 should clarify whether benefits in functional outcome and mortality can be achieved by early tracheostomy in ICU patients [1].

3 Study design

The Stroke-related Early Tracheostomy vs. Prolonged Orotracheal Intubation in Neurocritical care Trial 2 (SETPOINT2) is a multicenter, prospective, randomized, open, blinded endpoint (PROBE-design) trial. Patients with acute ischemic stroke, intracerebral haemorrhage or subarachnoid haemorrhage who are so severely affected that two weeks of ventilation are presumed necessary based on a prediction score are eligible. It was planned to enrol 190 patients per group (n = 380) [1]. To allow for early stopping in case of an overwhelming large treatment effect, a two stage group sequential design with one interim analysis was employed. The critical levels for the two-sided p-values for rejecting the null hypotheses assessed in confirmatory analysis were calculated according to O’Brien and Fleming [2]. The two sided significance level for early stopping for efficacy was $\alpha_1 = 0.0006$. In the final analysis, a significance level of $\alpha_2=0.0498$ is used to assure an overall two-sided type I error rate of $\alpha = 0.05$.

4 Analysis sets

Three analysis sets will be defined for the final analysis: the intention-to-treat (ITT) set, the per-protocol (PP) set and the safety analysis set. The intention-to-treat (ITT) set will include all randomized patients. Patients will be analysed as randomized according to the ITT principle. The per-protocol (PP) set includes all patients from the ITT set without major deviations from the study protocol. Major deviations from the study protocol are defined as

- not fulfilling the following inclusion criteria:
    - o confirmed admission diagnosis of AIS, ICH or SAH
    - o principal indication for tracheostomy
- fulfilling the following exclusion criteria:
    - o premorbid mRS > 1
    - o life-expectancy of less than 3 weeks
    - o ventilated more than 4 days
- SETscore < 8 at screening
- treatment not according to randomization (early tracheostomy performed later than 6 days after intubation or late tracheostomy performed earlier than 9 days after intubation)
- follow up time interview outside 180 days (6 month) +/- 14 days after admission to ICU
- missing mRS at 6 month after admission to ICU
- other major protocol violations (assessed by coordinating investigator before conduction of any analyses)

The safety analysis set consists of all randomized patients. Since an “as treated” assignment into two groups is not possible for the two interventions in this study, the analysis will be conducted using the following four groups:

1. Patients who were randomized to the early tracheostomy arm and who received the tracheostomy, died, or were extubated (extubation trial successful) on or before day 6 after intubation
2. Patients who were randomized to the early tracheostomy arm and who do not fulfil the criteria of the group above
3. Patients who were randomized to the late tracheostomy arm and who did not receive the tracheostomy before day 9 after intubation
4. Patients who were randomized to the late tracheostomy arm who did receive the tracheostomy before day 9 after intubation

Groups 1 and 3 represent patients that were treated as randomized (allowing for a time deviation of one day).

SAEs will be analysed in this safety set and in the ITT set. Since the predefined AEs are all tracheostomy related, the analysis of AEs will be conducted for a reduced safety set, where patients without a tracheostomy are excluded, and in an ITT set excluding all patients who did not receive a tracheostomy.

5 Definitions of endpoints to be analysed

5.1 Primary endpoint

The primary outcome is the dichotomized modified Rankin Score (mRS: A tool to describe functional outcomes after stroke) of 0-4 (good outcome) vs. 5-6 (poor outcome) at 6 months after admission to ICU. The mRS is a 7-point functional scale ranging from 0 (normal) to 6 (dead).

5.2 Secondary endpoints

All secondary endpoints are given in the following. In addition, some further endpoints not specified in the protocol are added and marked in italic.

Mortality

- overall mortality (death from any cause) until 6 months after admission to neurocritical care unit (NCCU)
- overall mortality during NCCU stay and in-hospital mortality
- neurologically caused death during NCCU (yes/no)
- cause of death during NCCU, in-hospital but after NCCU, within the 6 months follow-up period after discharge (free text fields)
- circumstance of death within the 6 months follow-up period after discharge:
    - sudden, unexpected death due to a stroke-related neurological condition,
    - sudden, unexpected death due to a stroke-related medical condition (e.g. aspiration pneumonia or a fall related to the stroke),
    - sudden, unexpected death due to an unrelated condition,
    - death following WLST (hospice or palliative care) due to stroke-related disability
    - death following WLST (hospice or palliative care) due to other medical conditions
- therapy withdrawal:
    - therapy withdrawal during NCCU stay (yes/no)
    - for subset of patients with therapy withdrawal: reason for withdrawal (categorized free text fields), time from NCCU admission until withdrawal

mRS

- *mRS dichotomized in 0-4 vs. 5-6 at time of NCCU discharge*
- mRS dichotomized in 0-3 vs. 4-6 *at time of NCCU discharge* and at 6 months after admission
- mRS assessed as a continuous variable *at time of NCCU discharge* and at 6 months after admission, as well as differences between *pre-hospital and NCCU discharge*, pre-hospital and 6 months after admission, and *NCCU discharge and 6 months after admission*

Ventilation and weaning

- time to first coma-free day (time from admission to NCCU until the first day, on which the patient is able to make sustained eye contact and/or follows commands)
- duration of mechanical ventilation (time from intubation until end of mechanical ventilation)
- time to first autonomous breathing (time from intubation until start of respirator weaning)
- duration of weaning (time from start of respirator weaning until end of mechanical ventilation)
- time from tracheostomy until decannulation
- orotracheal extubation trials:
    - at least one extubation trial
    - number of extubation trials
    - at least one successful extubation trial (>48 h)
    - number of successful extubation trials (>48 h)
    - reasons for not being extubated
    - reasons for not being successful
- successful weaning from mechanical ventilation

Intracranial pressure (ICP) increases > 25 mmHg

- occurrence of any ICP rise episodes before and after tracheostomy
- number of ICP rise episodes before and after tracheostomy

Sedation and medications

- number of days on which the following infusions (individually) were administered:
    - sedatives
    - opioids
    - vasopressors
    - antibiotics
- time to cessation of sedation (time from intubation until the day after the last day requiring sedatives)
- consciousness and sedation scores: RASS, SAS

Discharge and follow-up

- duration of NCCU stay (time from admission to discharge from NCCU)
- *discharge destination from NCCU (home, hospital, rehab-center, long-term care facility, other)*
- hospital length of stay (time from admission to discharge from hospital)
- *discharge destination from hospital (home, rehab-center, long-term care facility, other)*
- *location at 6 month follow-up (home, hospital, rehab-center, long-term care facility, other)*
- *requirement of a primary caregiver at 6 months after admission (yes, no, dead, loss-to-follow up)*

- caregiver assessment questionnaires at discharge from NCCU and at 6 months after admission
- Burden scale for Family caregivers (BSFC) at 6 months after admission
- patient reported outcome questions on the processes and results of care at 6 months after admission
- quality of life measured by EuroQol EQ-5D-5L at 6 months after admission (index score, items separately and EQ VAS)

5.3 Safety endpoints

The following **adverse events (AEs)** will be analysed as part of the safety analysis:

Periprocedural tracheostomy-related adverse events (during up to 2 hours after TT):

- Ventilation-related AEs
    - Relevant hypoxia during tracheostomy (SpO2 < 90%) requiring augmentation of ventilation (no, grade I, grade II)
    - Significant atelectasis requiring recruitment (no, grade I, grade II)
    - Pneumothorax (no, grade I, grade II)
    - Hematothorax (no, grade I, grade II)
- Bleeding AEs
    - Venous bleeding (no, grade I, grade II)
    - Arterial bleeding (no, grade I, grade II)
- Local trauma AEs
    - Puncture of the tracheal pars membranacea (no, grade I, grade II)
    - Dilatation of the tracheal pars membranacea (no, grade I, grade II)
    - Cannula misplacement (no, grade I, grade II)
    - Subcutaneous emphysema or pneumomediastinum (no, grade I, grade II)
    - Fracture of tracheal cartilage (no, grade I, grade II)
    - Damage to larynx or neighboring structures (no, grade I, grade II)
    - Accidental decannulation requiring reintubation (no, grade I, grade II)
- Cerebral compromise
    - ICP > 25 mmHg for > 5 minutes requiring treatment (no, grade I, grade II)
    - Neurological deterioration (> 4 points in NIHSS) (no, grade I, grade II)

Early tracheostomy-related adverse events (from 2 hours after TT to discharge from NCCU):

- Infection AEs
    - Local infection at tracheostomy site (no, grade I, grade II)
    - (Aspiration) pneumonia within first 48h post TT (no, grade I, grade II)
    - Mediastinitis (no, grade I, grade II)
- Tracheostomy tube problems
    - Cuff leak or rupture requiring change of cannula (no, grade I, grade II)
    - Patient discomfort (e.g. coughing, gagging) or malpositioning (e.g. cuff leak) requiring revision (no, grade I, grade II)

Late tracheostomy-related adverse events (at 6 months after admission / premature study termination):

- Recurrent / chronic infection at tracheostomy site (no, grade I, grade II)

- Scarring / disturbed wound healing at tracheostomy site (no, grade I, grade II)
- Tracheocutaneous fistula (no, grade I, grade II)
- Tracheal instability/tracheomalacia with respiratory insufficiency or disturbance of vocalization (no, grade I, grade II)
- Clinically relevant tracheal stenosis (no, grade I, grade II)
- Complicated change of cannula (no, grade I, grade II)
- Need for surgical revision of tracheal stoma (no, grade I, grade II)

In respect of **serious adverse events (SAEs),** the following parameters will be analysed:

- at least one serious adverse event during NCCU stay and during the whole observation period; all SAEs combined as well as stratified by
    - intensity
    - relation to tracheostomy
    - outcome
    - SAE categories (after categorization)
- number of SAEs per patient during NCCU stay and during the whole observation period

6 Data handling

6.1 Imputation of mRS

In case of patients lost to follow up before evaluation of the primary endpoint, missing data for the outcome will be imputed in the ITT set using multiple imputation as described by Allison [3] and van Buuren [4].

The fully conditional specification (FCS) method will be applied to impute missing values in the categorized modified Rankin Score (0-4 vs. 5 vs. 6) by means of an ordinal logistic regression model. This method is appropriate for an arbitrary structure of missing values which is the most general form of a missing data pattern (see, for example, Berglund [5]). The primary model fits the categorized modified Rankin Score 6 month after admission in ICU. In the imputation model, variables from the primary model (the outcome and the predictor variables group, age and Glasgow Coma Scale) will be included to take all information available into consideration for imputing missing data. After imputation, the categories 5 and 6 of the mRS score will be combined for the evaluation of the first hypothesis (categories 0-4 vs. 5-6), and categories 0-4 and 5 will be combined (categories 0-5 vs. 6) for the evaluation of the second hypothesis regarding the overall death rate (see Section 7.3 for the definition of the hypotheses).

The FCS method will also be applied to impute missing values in the Glasgow Coma Scale. The imputation model includes variables from the primary model (the mRS and the predictor variables group and age).

Fifty data sets will be created and the primary model will be fitted to each of them. After that, the obtained results will be combined using Rubin’s Rule [6]. The seed will be set to 202010 to be able to replicate the results. To implement the multiple imputation approach combined with the primary model, the procedures “mi”, “logistic”, and “mianalyze” within SAS (SAS Institut Inc., Cary, NC, USA) will be used.

Imputation will be done according to the following example (variable and data set names may have to be adapted):

```
proc mi data=ITT seed=202010 nimpute=50 out=ITT_imputed;
  class mrs_impute group;
  fcs logistic (mrs_impute= group age base_gcs / order=internal);
  var  group age base_gcs mrs_group;
run;
```

For the imputation of the other dichotomization (0-3 vs. 4-6) of the mRS 6 months after admission, the same methods will be applied modeling this dichotomization directly.

6.2 Missing data in the remaining endpoints

Missing values in secondary endpoints not described in Section 5.1 will not be imputed.

Due to an amendment to the study protocol, the eCRF was amended as well. Consequently, data for patients who were recruited before the amendment was recorded on the old version of the eCRF. Some baseline variables and endpoints were only recorded on the newer version of the eCRF, and thus these endpoints are missing in the pre-amendment collective of patients. There will be an indication in the final analysis for endpoints where this was the case.

The following variables were not recorded in the first version of the CRF:

Baseline

- race
- ethnicity
- APS score (of APACHE II score)

Secondary Endpoints

- mRS at NCCU discharge
- reason for death at 6 month follow-up
- reasons for extubation trial failures
- (daily) reasons for not being extubated
- number of days where medication is administered
- RASS and SAS scores
- destination of hospital discharge
- time to first coma-free day
- requirement of primary caregiver
- caregiver assessment questionnaires
- patient reported outcomes

7 Statistical methods

Statistical methods are used to describe the intervention groups at baseline, to evaluate differences between the control and experimental treatment for primary and secondary endpoints and to evaluate safety issues.

A CONSORT [7] flow diagram will be created to display the progress of all participants through the trial. This includes the number of randomized patients (per group), the number of patients

in the ITT and PP population, and reported reasons for exclusion from the ITT and from the PP set will be summarized per treatment group. Number of patients assessed for eligibility and the number of patients excluded because they did not meet inclusion criteria will be given for screened patients since protocol version 2.0 (information on patients screened before remained at the centres).

For the confirmatory analysis of the primary endpoint, the original SAS output will be reported to ensure transparency. In the following, analyses planned for baseline characteristics, the primary and secondary endpoints and safety endpoints are described in detail.

7.1 Descriptive methods

Continuous variables will be described using number of observations, mean, standard deviation, median, Q1, Q3, minimum, maximum and number of missing values. For categorical variables, absolute and relative frequencies will be given with missing values being reported as a separate category. Percentages for categorical variables will be based on all non-missing values in the respective groups. For categorical endpoints with subcategories, frequency tables as shown exemplarily in **Fehler! Verweisquelle konnte nicht gefunden werden.** will be provided. Time to event endpoints will be described by Kaplan-Maier curves (including the numbers at risk for different time points), median survival time (or other quantiles if median survival time cannot be calculated), median event time for all patients with an event, number of events and number of censorings. The descriptive methods described above will be used separately for each treatment group.

If a statistical test is conducted to compare treatment groups (as indicated below), point estimates with respective 95% confidence intervals will be provided additionally to the descriptive p-value.

Table 1: Example of frequency table with subcategories

	early tracheostomy N=50	prolonged intubation N=50	Overall N=100
Weaned from ventilation:			
-yes	**30 (60%)**	**40 (80%)**	**70 (70%)**
-after extubation	15 (30%)	20 (40%)	35 (35%)
-after tracheostomy	15 (30%)	20 (40%)	35 (35%)
-no	**20 (40%)**	**10 (20%)**	**30 (30%)**
-died during orotracheal ventilation	10 (20%)	5 (10%)	15 (15%)
-died after tracheostomy	5 (10%)	5 (10%)	10 (10%)
-permanent ventilation	5 (10%)	0 (0%)	5 (5%)

7.2 Baseline characteristics and description of therapy

The following baseline characteristics will be presented for the ITT, the PP and the safety set in the respective groups using the descriptive methods described above:

- age [years]
- gender
- race (American Indian/Alaska native, Asian, Black/African American, Hawaiian/Pacific Islander, White, Multi-race, other)
- ethnicity (Hispanic, Non-Hispanic)
- center (all centers separately as well as US-American and European centers combined)
- pre-hospital mRS (as categorical variable)
- Glasgow Come Scale (GCS) score
- NIH Stroke Scale (NIHSS) score
- APS Score (of APACHE II score)
- Lung Injury Score (LIS)
- SETscore including single items
- diagnosis (AIS, ICH, SAH) including diagnosis details

The following tracheostomy parameters will be described separately for each treatment group using descriptive methods described above:

- tracheostomy performed (yes/no)
- for subset of patients received tracheostomy: type of tracheostomy (percutaneous dilational TT or surgical TT), time from intubation to tracheostomy
- for subset of patients received no tracheostomy: reason why patient did not receive a tracheostomy (death, extubation, other)

Additionally, a frequency table with subcategories (where relative frequencies are based on all patients in the analysis set) will be presented for the following categories:

- received tracheostomy
    - o percutaneous dilational
    - o surgical
- did not receive tracheostomy
    - o extubated
    - o died
    - o other

7.3 Primary analysis

The primary aim of the trial is the comparison between early tracheostomy (experimental therapy) and ongoing orotracheal intubation group (control) with respect to the dichotomized modified Rankin Scale (mRS) score of 0-4 (treatment success) vs 5-6 (poor outcome) at 6 months after admission to ICU. Secondarily, in a hierarchical manner, the mRS score at 6 months after admission dichotomized into the groups 0-5 (=alive) and 6 (=dead) will be compared and analysed within a stagewise hierarchical testing procedure [8] described below

using the significance level of α_2=0.0498 in this final analysis due to the two stage group sequential design (as described above).

The multiple testing procedure for *a priori* hierarchically ordered hypotheses will be conducted as follows. Let $p_{S,E}$ and $p_{S,C}$ be the treatment success rate in the experimental and control group, and $p_{D,E}$ and $p_{D,C}$ be the death rate in the treatment and control group, respectively. At first, the hypothesis

$H_{0,S}$: $p_{S,E} = p_{S,C}$ vs. $H_{1,S}$: $p_{S,E} \neq p_{S,C}$

will be tested at level α_2. If $H_{0,S}$ can be rejected, the hypothesis

$H_{0,D}$: $p_{D,E} = p_{D,C}$ vs. $H_{1,D}$: $p_{D,E} \neq p_{D,C}$

will be tested at level α_2 within the confirmatory analysis. In case $H_{0,S}$ cannot be rejected, the procedure stops and $H_{0,D}$ is not tested in a confirmatory manner. This procedure together with the O'Brien Fleming levels controls the familywise error rate at level $\alpha = 0.05$.

Both endpoints will be analyzed using a logistic regression model including the factor treatment group and the covariates age, Glasgow Coma Scale at admission, and centre (US-American vs. European centres). The primary analysis will be conducted based on the ITT set using the imputed datasets (see Section 6.1).

The analysis will be performed according to the following SAS code (variable and data set names may have to be adapted):

```
proc logistic data = ITT_imputed;
  class mrs_group group(ref='0') centre_group / param=ref;
  model mrs_group(event='1')= group age base_gcs centre_group;
  by _imputation_;
  ods output ParameterEstimates=params;
run;

proc mianalyze parms=params;
  modeleffects group age base_gcs centre_group;
  ods output ParameterEstimates=params_mianalyze;
run;
```

As treatment effect estimate the unbiased median regression coefficient estimate for the group variable will be calculated using the SEQTEST procedure in SAS with MLE ordering, and the respective odds ratio together with the 95% confidence interval will be given. Moreover, combined effects for all other variables resulting from the logistic regression model will be given using odds ratios together with 95% confidence intervals.

7.4 Sensitivity analyses

A sensitivity analysis based on the PP set will be conducted using the same model specification as in the primary analysis. Furthermore, a complete case analysis for the ITT set will be conducted based on the same logistic model as in the primary analysis (without imputed values).

Moreover, the following subgroup analyses will be performed:

- age in categories <55, 55-65, >65 years
- male vs. female patients
- US vs. German centres
- lower enrolling vs. higher enrolling centres (<30 vs. >=30 randomized patients)
- primary diagnosis SAH, ICH, AIS
- Glasgow Coma Scale at admission (<6 vs. >=6)

For the analyses of treatment effects in subgroups, the primary analysis will be repeated for the respective subsets of patients. Therefore, the same model as in the primary analysis is used except for the analysis of US and German centres, where this variable has to be omitted in the model. Odds ratios together with 95% confidence intervals will be reported for the treatment effect in each subgroup. In addition, for all patients in the ITT set, the same model as in the primary analysis is repeated but the respective subgroup variable and an interaction of subgroup variable and treatment group is added as covariates. In this case, age as a continuous variable will be omitted for the analysis of the age subgroups, and Glasgow Coma Scale at admission for the analysis of subgroups defined by the Glasgow Coma Scale categorization. The p-value (only descriptive) of the interaction will be presented. Results will be summarized and presented using a forest plot.

7.5 Analysis of the secondary endpoints and further description

Analysis of secondary endpoints will be performed based on the ITT set and additionally, on the PP set as sensitivity analysis.

Mortality

Overall mortality until 6 months after admission will be analysed within the confirmatory analysis (see Section 7.3). In addition to the confirmatory analysis, the following analyses will be conducted: Categorical descriptive analysis will be performed on the mortality rate in the following time intervals: baseline to NCCU discharge, baseline to hospital discharge, baseline to 6 months follow-up. Descriptive survival analysis will be performed and Kaplan-Meier estimates for the survival rates at the 6 months follow-up time point will be provided. A log rank test and a cox regression model including the same covariates as in the primary analysis model will be used to compare both intervention groups regarding overall survival.

Causes of death entered in a free text field (during NCCU, in-hospital after NCCU, within 6 months follow-up after discharge) will be described via listings separated by treatment group.

Categorical descriptive analysis will be performed for neurologically caused deaths during NCCU (in the subgroup of patients who died during NCCU), for circumstances the patient died during follow-up (in the subgroup of patients who died during follow-up), and for therapy withdrawal during NCCU. Each endpoint will be compared using a chi-square test, respectively.

In the subgroup of mortality cases during NCCU stay, categorical descriptive analysis will be applied to the rate of mortalities following withdrawal of therapy, and a chi-square test will be performed.

Continuous descriptive analysis including a t-test will be applied to the time between NCCU admission and withdrawal of therapy.

mRS

Categorical descriptive analysis will be performed on the two different mRS score dichotomizations at the respective time points and a chi-square test will be performed to compare both treatment groups. Moreover, the logistic regression model specified for the primary analysis will be repeated for the alternative categorization including imputation.

Continuous and categorical (without dichotomization) descriptive analysis will be performed on the mRS score at the different time points as well as the differences between time points. A Mann-Whitney U test will be performed to compare both treatment groups.

Ventilation and weaning

Time to first coma-free day, duration of mechanical ventilation, time to first autonomous breathing, duration of weaning, and time from tracheostomy until decannulation will be described using Kaplan-Meier plots. If a patient didn't experience the event, the patient is censored at the end of follow-up of the patient. If a patient died before having the respective event, the patient is censored at the time when the last event in the whole population was observed or at 6 months (end of follow-up) if at least one patient is still alive and didn't experienced the event at the end of follow-up. A log-rank test will be used to compare both treatment groups.

Categorical descriptive analysis will be performed for receiving at least one orotracheal extubation trial, the number of extubation trials (0, 1, 2 or 3), at least one successful extubation trial and for the number of successful extubation trials including chi-square tests to compare both treatment groups, respectively. Moreover, a frequency table with subcategories will be presented as follows:

- Received at least one orotracheal extubation trial
    - Successfully extubated at least once
    - Not once successfully extubated
- Did not receive orotracheal extubation trial
    - Died
    - Received tracheostomy

Reasons for extubation trial failures will be listed separated by treatment group.

The following reasons for not being extubated on a NCCU day were prespecified in the eCRF:

- Unstable neurological condition
- Impaired respiratory automaticity
- Cardiopulmonary dysfunction
- Excessive secretions
- Inadequate airway protective reflexes
- Neuromuscular weakness
- Planned procedure
- Unable to safely reduce sedation
- Already tracheotomized
- Already spontaneously breathing

- Other

In the subset of NCCU days where the patient was still orotrachealy intubed (i.e. not “already tracheotomised” and not “already spontaneous[ly] breathing”), for each respective reason, continuous descriptive analysis will be performed on the number of NCCU days per patient where the reason was recorded. Moreover, for each respective reason it will be analyzed if this reason was present for a patient on at least one day using categorical descriptive analysis.

Categorical descriptive analysis will be applied to the rate of patients successfully weaned from mechanical ventilation. Subcategories will be allocated as follows:

- Successfully weaned
    - Weaned from orotracheal ventilation (extubated)
    - Weaned from ventilation after tracheostomy
- Not successfully weaned
    - Died during orotracheal ventilation
    - Died after tracheostomy
    - Permanent ventilation (end of mechanical ventilation = no and not dead)

A chi-square test will be applied to evaluate the difference between successful and unsuccessful ventilator weaning between treatment groups.

Intracranial pressure (ICP) increases >25 mmHg

Categorical descriptive analysis including a chi-square test will be performed on the proportion of patients with ICP increases >25 mmHg. Subcategories will be allocated as follows:

- ICP increase >25 mmHg
    - Only before tracheostomy
    - Only after tracheostomy
    - Before and after tracheostomy
- No ICP increase >25 mmHg
- Not measured

Continuous descriptive analysis including a t-test will be performed on the number of ICP increases >25 mmHg at any time, and separately for before and after tracheostomy. This will be conducted for the absolute numbers as well as for the relative numbers to the number of days on NCCU, number of days before tracheostomy and number of days on NCCU after tracheostomy, respectively.

Sedation and medications

Continuous descriptive analysis will be performed on the number of NCCU days where each of the respective medications was administered and compared using a t-test.

Time to cessation of sedation will be described using Kaplan-Meier plot. Failure to observe the event (cessation of sedation), or death before observing the event will be right-censored at NCCU day 28. A log-rank test will be used to compare both treatment groups.

The median RASS and SAS score per patient during NCCU stay will be calculated. Then, categorical descriptive analysis will be performed on the median per patient RASS and SAS scores, and compared using a Mann-Whitney-U-test.

Discharge and follow-up

Continuous descriptive analysis will be performed on the NCCU and on the hospital length of stay for all patients. Durations will be compared between both intervention groups using a t-test. This comparison serves as a proxy for costs and should not be considered for an assessment of the treatment effect, because early death leads to shorter hospital stays.

In addition, to correct for early deaths, time to NCCU discharge and time to hospital discharge will be described using Kaplan-Meier plots. If a patient dies before discharge, the patient is censored at the time when the last discharge in the whole population was observed or at 6 months (end of follow-up) if at least one patient is still at NCCU or hospital at the end of follow-up. A log-rank test will be used to compare both treatment groups.

Categorical descriptive analysis as well as a chi-square test will be performed on the discharge destinations from NCCU (home, hospital, rehab-center, long-term care facility, other), from hospital (home, rehab-center, long-term care facility, other) and for the location at the 6 month follow-up (home, hospital, rehab-center, long-term care facility, other).

Categorical descriptive analysis including a chi-square test will be performed on requirement of a primary caregiver at the 6 months follow-up.

Categorical descriptive analysis will be performed on the six primary caregiver questions at NCCU discharge and on the two primary caregiver questions at the 6 months follow-up. A chi-square test will be used to compare both groups.

The Burden Scale for Family caregivers (BSFC) score will be calculated according to Figure 1. The score will be categorized according to Figure 2. Categorical description will be performed on the burden score categorizations (none to mild, moderate, severe to very severe). In addition, continuous descriptive analysis and a Mann-Whitney-U-test will be conducted for the uncategorized score.

Patient reported outcomes will be analysed using categorical descriptive analysis on the reported suffering from wheezing, neck or throat pain, difficulty breathing, weak or impaired speech due to tracheal injury, or other problems related to tracheotomy, as well as on the responses to the other three patient reported outcome questions, respectively. Groups will be compared using a chi-square test, respectively.

The index score of the EQ-5D-5L questionnaire will be calculated according to the method described in Section 4.1 of [8], using the country specific values set for Germany [9] and the USA [10]: “An EQ-5D summary index is derived by applying a formula that attaches values (weights) to each of the levels in each dimension. The index is calculated by deducting the appropriate weights from 1, the value for full health (i.e. state 11111). The collection of index values (weights) for all possible EQ-5D health states is called a value set“ [9]. Scores will be set to missing, if at least one answer is missing. In the supplementary material 3 of [9], a database with all possible health states and the corresponding summary scores according to the German value set is provided. The corresponding database is not included in the publication of the US value set, but can easily be calculated by the algorithm described above using the US specific parameters. After calculating the health state for each patient, these

databases will be used to assign a summary score to each patient. The EQ-5D-5L index value and VAS scores then will be analysed according to the recommendation of the manual [8]. The analysis strategy is described in the following. Specifically, continuous descriptive analysis will be performed for the index value and VAS scores, separated by treatment group. A histogram, visualizing the distribution of the VAS scores similar to Figure 3: **EQ-5D-5L VAS distribution** (c.f. [8]) will be created. As in Figure 3, the histogram will have a bin-width of 6.67. A Mann-Whitney-U test for the EQ-5D-5L index score is used for comparing both treatment groups.

Categorical descriptive analysis for each item of the questionnaire will be performed.

SHORT TEST: Burden of family caregivers

Short version of the Burden Scale for Family Caregivers by Prof. Dr. Elmar Gräßel

We are asking you for information about your present situation. The present situation comprises the caregiving you provide because of the illness of your family member (or friend). The following statements often refer to the type of assistance you provide. This may be any kind of support up to nursing care.

Please draw an "X" for the best description of your present situation.
Please answer every question!

	strongly agree	**agree**	**disagree**	**strongly disagree**	
My life satisfaction has suffered because of the care.					
I often feel physically exhausted.					
From time to time, I wish I could "run away" from the situation I am in.					
Sometimes I don't really feel like "myself" as I did before.					
Since I have been a caregiver, my financial situation has decreased.					
My health is affected by the care situation.					
The care takes a lot of my own strength.					
I feel torn between the demands of my environment (such as family) and the demands of the care.					
I am worried about my future because of the care I give.					
My relationships with other family members, relatives, friends, and acquaintances are suffering as a result of the care.					**Points reached (Sum score)**
	[number of marks] x3	[number of marks] x2	[number of marks] x1	[number of marks] x0	

Fig. 2 Example of an information brochure. Scale included for the informal caregiver to fill out without assistance

Figure 1: Burden scale calculation [11]

Evaluation of the number of points:

0 - 4 points **Your burden of care is: none to mild**

It means:
You do not have an increased risk of physical discomfort that is above the usual level of complaints in your age group.
Recommendation:
Support the family caregiver to the extent that he/she continues to do well.

5 - 14 points **Your burden of care is: moderate**

It means:
You have an increased risk of physical discomfort that is above the usual level of complaints in your age group.
Recommendation:
Try to relieve yourself and get more rest. You should have your physical health evaluated.

15 - 30 points **Your burden of care is: severe to very severe**

It means:
You have a greatly increased risk of physical discomfort that is above the usual level of discomfort in your age group.
Recommendation:
Take steps to reduce the extent of your physical discomfort. It might help to get some relief by finding others who can support the care receiver as well. A health check is urgently recommended in your case.

Fig. 3 Support for interpreting the results and concrete recommendations to be given to the informal caregiver

Figure 2: Burden scale categorization [11]

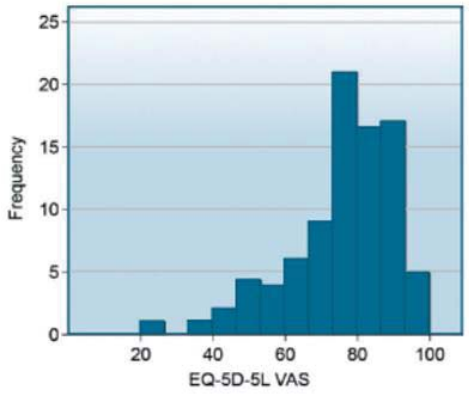

Figure 3: EQ-5D-5L VAS distribution [8]

7.6 Safety analyses

Adverse events

All tracheostomy related adverse events will be analysed in the reduced ITT set and in the reduced safety set (see Section 4). For all AEs listed in Section 5.3, categorical descriptive analysis will be performed with categories no, Grade I, Grade II. Additionally, categorical descriptive analysis will be performed on the respective rates of patients that suffered any (Grade I or Grade II) ventilation related AE, any bleeding AE, any local trauma AE, any cerebral

compromise, any infection or any tracheostomy tube problems, respectively. For each comparison a descriptive p-value using a chi-square test will be given. AEs may be visualized using appropriate graphics.

Serious adverse events

Serious adverse events will be analysed separately for both groups defined in the ITT set and for the four groups of the safety set. Categorical descriptive analysis and a chi-square test will be performed on the rate of patients who suffered at least one serious adverse event during NCCU stay and during the whole observation period. This will be conducted for all SAEs combined and stratified by intensity and relation to tracheostomy. Furthermore, categorical descriptive analysis will be performed on the number of SAEs per patient.

All serious adverse events will be listed together with the intensity, relation to tracheostomy, result and time period where the SAE occurred (during NCCU, after discharge from NCCU) separately for both treatment groups.

8 Interpretation of results

Significant p-values in the logistic regression model in both of the hierarchically ordered testing problems ($H_{0,S}$ and $H_{0,D}$) let us conclude that the odds of treatment success and the odds of death in the experimental group are different from the corresponding odds in the control group when adjusting for age, Glasgow Coma Scale at admission and centre in the follow sense: The ratios of the respective odds are different from 1 with a type-1 error probability of 5%.

Given our data and model, the given point estimates for the odds ratios are the most likely ones (“maximum likelihood estimates”).

A significant p-values for the hypothesis test associated with $H_{0,S}$, but an insignificant p-value for the test associated with $H_{0,D}$, let us conclude that the odds of treatment success in the experimental group are different from the odds of treatment success in the control group in the sense that their ratio is different from 1 with a type-1 error probability of 5%.

Due to the hierarchical nature of the testing problem, none of the null-hypotheses may be rejected if we observe no significant p-value in the testing problem associated with $H_{0,S}$ irrespective of the resulting p-value in the testing problem associated with $H_{0,D}$.

9 Software

SAS version 9.4 or higher will be used for all analyses.

10 References

[1] S. Schönenberger, W.-D. Niesen, H. Fuhrer, C. Bauza, C. Klose, M. Kieser, J. I. Suarez, D. B. Seder, J. Bösel, "Early tracheostomy in ventilated stroke patients: Study protocol of the international multicentre randomized trial SETPOINT2 (Stroke-related Early Tracheostomy vs. Prolonged Orotracheal Intubation in Neurocritical care Trial 2)," *Int. J. Stroke*, 2016, doi: 10.1177/1747493015616638.

[2] P. C. O'Brien and T. R. Fleming, "A Multiple Testing Procedure for Clinical Trials," *Biometrics*, 1979, doi: 10.2307/2530245.

[3] P. D. Allison, *Missing Data.* Sage University Papers Series *on Quantitative Applications in the Social Sciences*.07-136, Thousand Oaks, CA: Sage, 2001.

[4] S. van Buuren, *Flexible Imputation of Missing Data*, Second Edition. New York: Chapman and Hall/CRC, 2018.

[5] P. A. Berglund, "An Introduction to Multiple Imputation of Complex Sample Data using SAS v.9.2 (Paper 265-2010)," 2010.

[6] W. M. Campion and D. B. Rubin, "Multiple Imputation for Nonresponse in Surveys," *J. Mark. Res.*, 1989, doi: 10.2307/3172772.

[7] K. F. Schulz, D. G. Altman, and D. Moher, "CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials," *BMJ*, vol. 340, no. 7748, pp. 698–702, 2010, doi: 10.1136/bmj.c332.

[8] R. Rabin, M. Oemar, M. Oppe, B. Janssen, and M. Herdman, "EQ-5D-5L user guide," *Basic Inf. how to use EQ-5D-5L Instrum.*, 2015.

[9] K. Ludwig, J. M. Graf von der Schulenburg, and W. Greiner, "German Value Set for the EQ-5D-5L," *Pharmacoeconomics*, vol. 36, no. 6, pp. 663–674, 2018, doi: 10.1007/s40273-018-0615-8.

[10] A. S. Pickard *et al.*, "United States Valuation of EQ-5D-5L Health States Using an International Protocol," *Value Heal.*, vol. 22, no. 8, pp. 931–941, 2019, doi: 10.1016/j.jval.2019.02.009.

[11] A. Pendergrass, C. Malnis, U. Graf, S. Engel, and E. Graessel, "Screening for caregivers at risk: Extended validation of the short version of the Burden Scale for Family Caregivers (BSFC-s) with a valid classification system for caregivers caring for an older person at home," *BMC Health Serv. Res.*, 2018, doi: 10.1186/s12913-018-3047-4.