FERRING
PHARMACEUTICALS

Statistical Analysis Plan

Title of trial: A Double-blind, Randomised, Placebo-controlled, Phase 2b/3 Adaptive Clinical Trial Investigating the Efficacy and Safety of Selepressin as Treatment for Patients with Vasopressor-dependent Septic Shock
NCT number: NCT02508649
Sponsor trial code: 000133
Date: 20 Nov 2017

STATISTICAL ANALYSIS PLAN

A Double-blind, Randomised, Placebo-controlled Phase 2b/3 Adaptive Clinical Trial Investigating the Efficacy and Safety of Selepressin as Treatment for Patients with Vasopressor-dependent Septic Shock

SEPSIS-ACT

Selepressin Evaluation Programme for Sepsis-Induced Shock - Adaptive Clinical Trial

000133

Investigational Product:	Selepressin; concentrate for solution for infusion Placebo; sterile 0.9% sodium chloride solution
Indication:	Vasopressor-dependent Septic Shock
Phase:	2b/3
Author:	[redacted]
Date of issue:	November 20th - 2017
Version:	Final - Ver. 9.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes
1	Dec 17-2014	Original SAP	Not applicable
2	Dec 19-2014	Format change only. No content was changed	Version 1
3	Dec 19-2014	Format change only. No content was changed	Version 2
4	Mar 24-2015	Added endpoint: Episodes of hypotension Modified endpoint: Norepinephrine/ noradrenaline and other vasopressor doses Updated section on definition of P&VFDs and imputations for P&VFDs Various minor elaborations of endpoint definitions and minor changes to analyses	Version 3
5	May 27-2015	Added sensitivity analysis for primary endpoint. Appendix 4 notation updated and estimate for number of new organ dysfunctions and failures included Appendix 5 updated to reflect tenths of days. Appendix 6 updated to make proof more readable. Various editorial clarifications	Version 4

6	July 8 - 2016	Expanded definition of onset of shock to include ANY vasopressors Replaced two AE tables regarding adverse events based on changes in vital signs/laboratory values assessed as unanticipated in the setting of septic shock, as we do not collect sufficient data to produce these tables. Modified tables presented. The analysis of other vasopressors has been split into several components. Sensitivity analyses added to ICU-free days and ICU length of stay. Fluid balance/Urinary output analyses updated Terlipressin added to list of vasopressors for derivation of primary endpoint etc. Definition of primary endpoint imputation for patients lost to follow-up is extended to include any type of withdrawal before Day 30. List of critical adverse events updated (same as before, just re-grouped). No AE tables on AEs leading to withdrawal, as it is not an option to withdraw patient from trial due to AE. Added alternative analysis to “duration of” endpoints, in case normality cannot be assumed for the whole of the data. Small edit in formula in Appendix 5. Continuous version of baseline norepinephrine version is adjusted for weight (categorical version from strata is unadjusted) Added imputation rules for NE and other vasopressors (section 9.4.5) and IMP (section 8.1.1) Appendix 2 slightly updated with corrections. List of protocol deviations modified (section 5)	Version 5

7	Nov 9 - 2017	Additional analyses to MAP analyses. Analysis for duration of septic shock changed from ANCOVA to negative binomial (as blinded data suggests a better fit) with permutation test as backup in case model assumptions do not hold. Analyses for duration of ventilation and duration of RRT changed from ANCOVA to zero inflated negative binomial (as blinded data suggests a better fit) also with permutation test as backup in case model assumptions do not hold. Analysis for ICU length of stay and hospital length of stay changed from ANCOVA to permutation test (as blinded data suggested a non-parametric model). Covariates were dichotomized accordingly. Appendix 4 removed as no longer needed. Appendix 3 updated with variance correction for difference in PVFDs Analyses in 9.4.8, 9.4.9, and 9.4.10 will not be repeated for the selected arm only and data from part 2 only, as they are measured on a subset of patients, probably only in part 1. A “time” by “baseline value” interaction term has been added to the repeated ANCOVA analyses in 9.3.1.5, 9.3.3.1, 9.3.4.1, 9.3.4.2, and 9.4.5 as it is likely that the trajectory over time for each patient will depend on the baseline value. The supportive competing risk and Kaplan-Meier graphs for duration of RRT has been taken out as it makes no sense since RRT durations are short and repetitive. It is more meaningful to just look at 90 day mortality. For the remaining ‘duration’ endpoints withdrawal rates has been added to the competing risk and Kaplan-Meier graphs. As this is a superiority trial, the PP analysis will only be performed for the primary endpoint, and hence not performed for secondary as previously stated.	Version 6

		Only major protocol deviations will be listed. Adverse events graphs added to output. Fluid balance analyses updated. Analyses for norepinephrine/noradrenaline and other vasopressor doses updated. Forest plots added for incidence of RRT (section 9.3.1.8) and mortality (section 9.3.2.3).	
8	Nov 9 - 2017	Format change only. No content was changed	Version 7
9	Nov 20 - 2017	Format change only. No content was changed	Version 8

Signed agreement on Statistical Analysis Plan

The original analysis plan was reviewed by,

- [redacted] Global Biometrics, Ferring Pharmaceuticals A/S
- [redacted] Clinical R&D, Selepressin, Ferring Pharmaceuticals A/S
- [redacted] Global Pharmacovigilance, Ferring Pharmaceuticals A/S
- [redacted] Clinical R&D, Selepressin, Ferring Pharmaceuticals A/S
- [redacted] Health Economics & Outcomes Research, Clinical and Non-Clinical Development, Ferring Pharmaceuticals A/S

And reviewed and approved (signed electronically) by,

- [redacted] Biometrics, Ferring Pharmaceuticals A/S

Subsequent editions reviewed and approved (signed electronically) by,

- [redacted] Biometrics, Ferring Pharmaceuticals A/S

TABLE OF CONTENTS

1 Introduction

This document describes the planned statistical analyses for selepressin (FE202158) 000133, and is based on protocol dated July 8th 2016 including amendments 1 and 2.

All analyses in the SAP were planned and pre-specified prior to trial termination due to futility.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Terms	Definitions
Evaluable patient	Patient who has been treated with IMP and 30 days have passed since initiation of IMP infusion
Randomised	Patient randomised to trial treatment
Screened	Patient who enters the screening phase
Selepressin	FE 202158
Treated patient	Randomised and dosed
1-KM curve	1 minus Kaplan Meier curve

1.1.2 Abbreviations

Abbreviations	Meaning of abbreviations in document
AE	Adverse Event
ANCOVA	Analysis of covariance
AUC	Area Under the Curve
CVP	Central venous pressure
EVLW	Extra-vascular lung water
EQ-5D-5L	EuroQol – 5 Dimensions – 5 Levels
FAS	Full-Analysis Set
ICU	Intensive care unit
IMP	Investigational Medicinal Product
ITT	Intention-to-treat
LOCF	Last observation carried forward
MAP	Mean Arterial Pressure
MedDRA	Medical Dictionary for Regulatory Activities
NE	Norepinephrine
PK	Pharmacokinetic
PP	Per-Protocol
PPI	Pulmonary permeability index
PT	Preferred term
P&VFD	Vasopressor and mechanical ventilator free days
QALY	Quality adjusted life years
RAR	Response adaptive randomisation
RRT	Renal replacement therapy
ScvO2	Oxygen Saturation in Vena Cava Superior
SOFA	Sequential Organ Failure Assessment score

Abbreviations	Meaning of abbreviations in document
SOC	System Organ Class
WBC	White blood cells

2 Trial Objectives and Endpoints

2.1 Objectives

Primary Objective

- To demonstrate superiority of selepressin plus standard care versus placebo plus standard care in the number of vasopressor- and mechanical ventilator-free days (with penalty for mortality) in patients with vasopressor-dependent septic shock

Secondary Objectives

- To determine the efficacy of selepressin on:
  - Organ dysfunction
  - Morbidity and mortality
  - Fluid balance
  - Health-related quality of life
- To determine the safety profile of selepressin
- To determine the pharmacokinetics of selepressin
- To determine the health economics of selepressin
- To further evaluate a range of biomarkers in relation to the mode of action of selepressin

2.2 Endpoints

2.2.1 Primary Endpoint

- Vasopressor- and mechanical ventilator-free days (P&VFDs) up to day 30

2.2.2 Key Secondary Endpoints

- All-cause mortality (defined as the fraction of patients that have died, regardless of cause, by the end of Day 90)
- Renal replacement therapy (RRT)-free days up to Day 30 (excluding patients on RRT for chronic renal failure at time of randomisation)
- Intensive care unit (ICU)-free days up to Day 30

2.2.3 Secondary Endpoints

Organ dysfunction

- Vasopressor-free days up to Day 30
- Mechanical ventilator-free days up to Day 30
- Duration of septic shock up to Day 30

- Duration of mechanical ventilation up to Day 30
- Incidence of RRT up to Day 30 (counting patients who die as on RRT and excluding patients on RRT for chronic renal failure at time of randomisation)
- Duration of RRT up to Day 90 (excluding patients on RRT for chronic renal failure at time of randomisation)
- Daily overall and individual organ (cardiovascular, respiratory, renal, hepatic, coagulation) scores using the modified Sequential Organ Failure Assessment (SOFA) scores until ICU discharge
- Incidence of new organ dysfunctions and new organ failures (based on the SOFA score) up to Days 7 and 30

Morbidity/mortality

- ICU length of stay up to Day 30
- All-cause mortality (defined as the fraction of patients that have died, regardless of cause, by the end of Days 30 and 180)

Fluid balance

- Daily and cumulative fluid balance until ICU discharge (for a maximum of 7 days)
- Daily and cumulative urinary output until ICU discharge (for a maximum of 7 days)

Health-related quality of life

- Change in utility, based on EQ-5D-5L, up to Day 180

2.2.4 Safety Endpoints

- Incidence of adverse events (type, frequency, and intensity) with specific emphasis on:
  - Ischaemic events (e.g. myocardial, skin, cerebral, mesenteric, and limb ischaemia)
- Changes in vital signs and safety laboratory variables, including:
  - Number of clinically significant results assessed as unanticipated in the setting of septic shock
- Episodes of hypotension

2.2.5 Additional Endpoints

- Hospital-free days up to Day 90
- Hospital length of stay up to Day 90
- Patient residence at Day 30, Day 60, Day 90, and Day 180
- Mean arterial pressure (MAP), until ICU discharge (for a maximum of 7 days)

- Norepinephrine/noradrenaline and other vasopressor doses
- Pharmacokinetic response (in a subset of approximately 200 patients) - to be reported separately according to a pre-specified pharmacokinetic analysis plan
- Health economic evaluation – to be reported separately according to a pre-specified health economic analytical plan
- Creatinine Clearance
- PaO2/FiO2 ratio (in a subset of 100-350 patients)
- Extravascular Lung Water and Pulmonary Permeability Index (in a subset of 100-350 patients)
- Cardiac output (in a subset of 100-350 patients)
- Cytokines (in a subset of 100-350 patients)
- Angiopoietin 1 and 2 levels (in a subset of 100-350 patients)

2.2.6 Other Assessments

- Central Venous Pressure
- Central Venous Oxygen Saturation
- Arterial Blood Gases (PaO2, PaCO2, SaO2, pH, HCO3, base excess) and Lactate

3 Trial design

3.1 General Design Considerations

The overall adaptive design is a Phase 2b/3 trial, in which dose-ranging with response-adaptive randomization (RAR) (see Appendix 5 for details) is utilized in a first part (the Phase 2b part – Part 1), followed by a traditional 1:1 randomised comparison of selepressin to placebo in the second part (the Phase 3 part – Part 2). The final analysis uses patients from both parts of the trial. The entire trial, combining both parts, represents an adequate and well-controlled comparison of selepressin and placebo.

In Part 1 of the trial, up to four dosing regimens will be investigated.
Arm 1: Starting dose at 1.7 ng/kg/min, and a max. dose of 2.5 ng/kg/min
Arm 2: Starting dose at 2.5 ng/kg/min, and a max. dose of 3.75 ng/kg/min
Arm 3: Starting dose at 3.5 ng/kg/min, and a max. dose of 5.25 ng/kg/min
Arm 4: Starting dose at 5 ng/kg/min, and a max. dose of 7.5 ng/kg/min

Part 1 comprises a minimum of 300 evaluable patients and a maximum of 800 treated patients. During Part 1, patients will be randomised to placebo or selepressin (Arms 1 to 3). Arm 4 will only be opened between 200 evaluable- 600 treated patients and if there is at least a 50% probability that Arm 3 has a higher expected P&VFD than Arm 2 and if data from the lower dosing levels do not suggest any significant safety signals. To minimize the risk of imbalance between treatment arms, randomisation will be stratified based on trial site, the need for mechanical ventilation (Yes/No), norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and creatinine (< or ≥150 µmol/L) (See Appendix 7 for details).

Part 1 will begin with a 200-patients treated “burn-in” period during which fixed randomisation across the treatment arms will be used (one-third of the patients randomised to placebo and two-ninths of the patients to each of the selepressin arms [Arms 1 to 3]). The factors described above will be used to stratify the randomisation.

After completion of the burn-in period, Part 1 will utilize response-adaptive randomisation to preferentially place patients into the arms that appear to have the maximum benefit with respect to the primary endpoint. A fixed fraction (one third) of patients will be randomised to placebo throughout Part 1 to ensure contemporaneous control patients are enrolled throughout the trial.

If Part 1 culminates in the decision to run Part 2, Part 2 will be a 1:1 comparison of placebo to the best-performing active treatment arm. The best-performing active treatment arm will be identified at the end of Part 1. Part 2 will utilize a fixed 1:1 randomisation proportion, with stratified randomisation as described for Part 1. Part 2 can begin after any interim analysis after 300 evaluable -800 treated patients in Part 1 and the size of Part 2 will include enough patients to bring the total number of evaluable patients in Part 1 and Part 2 up to 1800, ensuring a minimum sample size of Part 2 of 1000 evaluable patients.

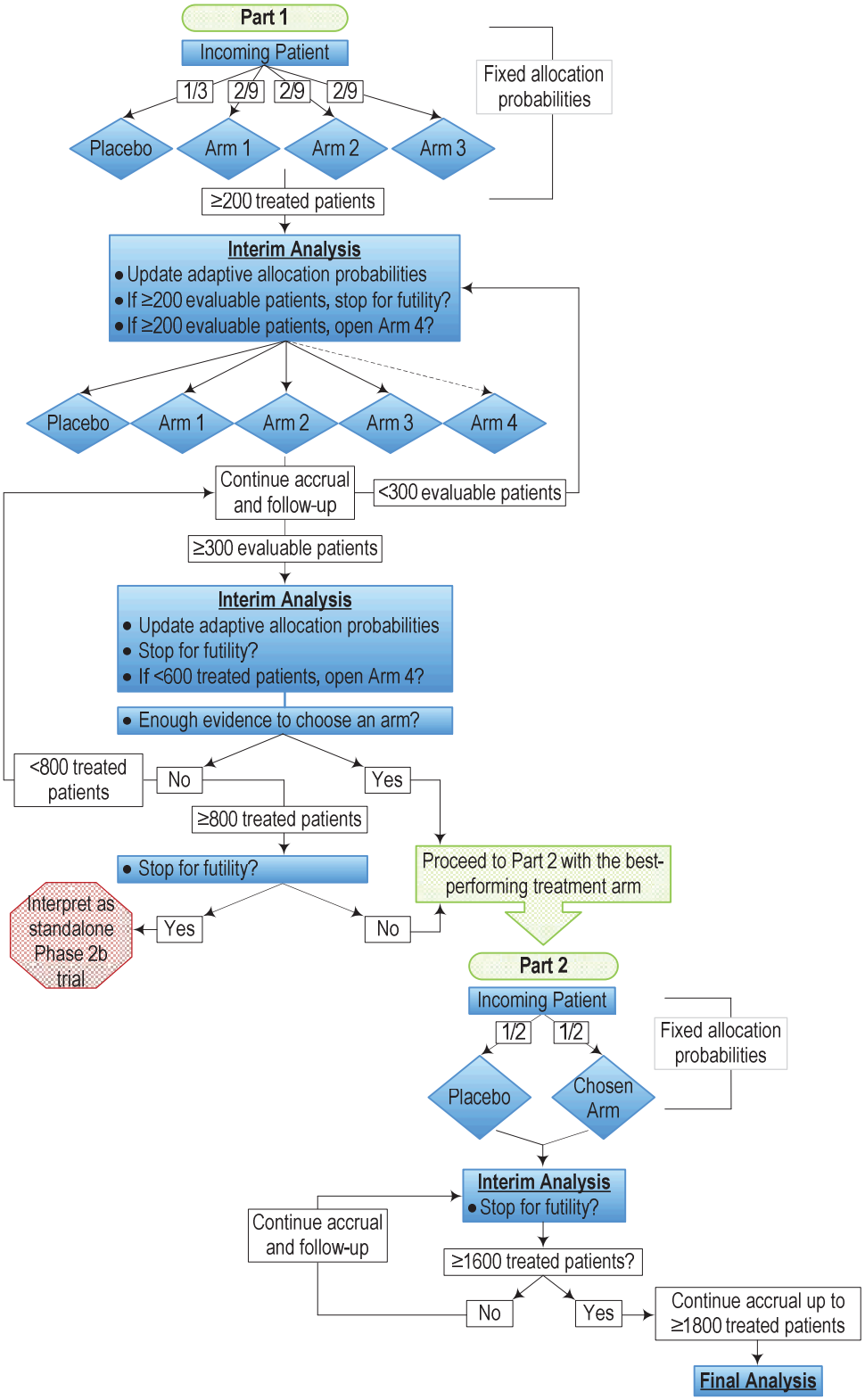

Figure 1: Design Flow Chart

3.2 Determination of Sample Size

At least 1800 evaluable patients combined for Part 1 and Part 2 are needed for the final analysis. The overall power of obtaining statistical significance based on combined evidence from Part 1 and Part 2 is 91% in situations where all 4 arms have a true underlying 1.5% lower mortality rate and a 1.5-day higher expected number of P&VFDs for survivors as compared to placebo (corresponding to an overall treatment effect of 1.5 P&VFDs). If the effect sizes are 2% on mortality and 2 days for P&VFDs in survivors for all 4 arms (corresponding to an overall treatment effect of 2 P&VFDs), then the overall power is 98%. In this latter case the probability of engaging into Part 2 is ~99%.

4 Patient Disposition

A summary table will present, for each part of the trial and overall, the number of patients in the population sets: 'Screened', 'Intention to treat', Full analysis set', 'Per protocol', 'Safety', 'Completed trial', 'Withdrawals', and 'IMP discontinuations' with a breakdown of reasons/categories for trial withdrawals and IMP discontinuations.

The patient disposition table will be broken down by each of the stratification variables [the need for mechanical ventilation (Yes/No), norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and creatinine (< or ≥150 µmol/L)] and broken down chronologically displaying number of patients 'completed' and 'withdrawn from trial' at Day 30, Day 90 and Day 180.

The number of patients screened but not randomised/allocated to treatment will be presented with the reason(s) for screen failure in a data listing.

All major protocol violations (including misrandomisations), based on the Full analysis set, will be summarised for each part of the trial.

Furthermore 1-KM plots, based on the ITT, will be presented for the time to trial withdrawals/IMP discontinuations (whichever comes first) differentiated by reason of trial withdrawal/IMP discontinuation using cumulative incidence functions. Dropout rates between treatment groups will be evaluated by the log-rank test.

5 Protocol Deviations

Patient data in the full analysis set (FAS) will be excluded from the per-protocol (PP) analysis set if they meet any of the following criteria:

- Dosing errors expected to significantly impact efficacy
- More than 16 hours from onset of vasopressor treatment to start of IMP
- Failed to receive a continuous infusion of norepinephrine/noradrenaline base greater than 4.5 ug/min for at least one hour or received less than 4.5 ug/min of norepinephrine/noradrenaline base at start of IMP infusion
- Violation of exclusion criteria 2: Primary cause of hypotension not due to sepsis (e.g., major trauma including traumatic brain injury, hemorrhage, burns, or congestive heart failure/cardiogenic shock)
- Violation of exclusion criteria 6: Chronic mechanical ventilation for any reason OR severe chronic obstructive pulmonary disease (COPD) requiring either continuous daily oxygen use during the preceding 30 days or mechanical ventilation (for acute exacerbation of COPD) during the preceding 30 days
- Violation of exclusion criteria 9: Decision to limit full care taken before obtaining informed consent
- Violation of exclusion criteria 10: Use of vasopressin in the past 12 hours prior to start of IMP infusion or use of terlipressin within 7 days prior to start of IMP infusion
- Violation of exclusion criteria 12: Prior use of an investigational medicinal product within the last month OR planned or concurrent participation in a clinical trial for any investigational drug or investigational device

Furthermore, any other major protocol violations, such as serious unforeseen violations deemed to invalidate the data and affect the conclusions of the study will lead to exclusion of data from the PP analysis set.

Major protocol deviations will lead to exclusion of data from the PP analysis, while data will not be excluded because of minor protocol deviations. The list of major protocol deviations will be detailed and documented in the clean file document prior to database release.

All major protocol deviations will be listed in patient data listings.

6 Analysis sets

6.1 Intention-To-Treat Analysis Set

The intention-to-treat (ITT) analysis set comprises of all randomised (as planned) patients.

6.2 Full-Analysis Set

The FAS comprises data from all randomised (as planned) and dosed patients.

6.3 Per Protocol Analysis Set

Patients in the FAS will be excluded from the PP analysis set if they meet any major protocol violations defined in (Section 5). Data will be used up to the point of protocol violation.

6.4 Safety Analysis Set

The safety analysis set comprises all treated patients and are analysed according to the actual treatment received.

7 Trial population

7.1 Demographics and Other Baseline Characteristics

Categorical data will be summarised using numbers and percentages. The percentages are based on the total number of patients with a corresponding assessment. Continuous data will be presented, for example, using the number of patients (N), mean and standard deviation, median, interquartile range, minimum and maximum. All baseline characteristics will be listed.

Demographics and baseline characteristics of the study population will be summarised for the FAS.

7.1.1 Demographics

Descriptive statistics of baseline demographics variables will be summarized by treatment arm and total.

7.1.2 Vital Signs at Baseline

Baseline vital signs will be summarised by treatment arm and total.

7.1.3 SOFA Score, APACHE II Score and Septic Shock Characteristics

Baseline SOFA score (modified), APACHE II score and information on septic shock (infection proven/suspected, primary infection type and location will be summarised by treatment arm and total.

7.2 Medical History

Medical history recorded at screening visit will be summarised by treatment arm and total.

Furthermore, medical history will be presented in patient data listings.

7.3 Prior and Concomitant Medication

Prior and concomitant medication will be summarised by treatment arm and total.

Furthermore, concomitant medication will be presented in patient data listings.

8 Exposure and Treatment Compliance

8.1.1 Extent of Exposure

The total amount (adjusted by weight (µg/kg)) of selepressin administered and the number of days (reported to one decimal place) treated with selepressin will be summarised by (active) treatment arm and total (active treatment arms).

Furthermore, the mean cumulative amount administered and the mean infusion rate will be tabulated by treatment arm and presented graphically (also by treatment arm and total).

If a patient has missing infusion rate and the patient is still in the trial (not dead or withdrawn) it will be assumed that selepressin was not administered and a value of zero will be imputed, unless there is an interval in the timing log covering the exact time point (8 AM and 8 PM is the assumed time point for missing morning and evening collection time points). In that case LOCF will be used, but only within the time interval.

If a patients has missing cumulative selepressin volume and the patient is still in the trial (not dead or withdrawn), LOCF will be used assuming that selepressin was not administered and hence keeping the cumulative volume constant.

9 Efficacy

9.1 General Considerations

All statistical tests will be performed using a two-sided test at a 5% significance level.

If the trial is stopped prematurely due to e.g. futility, the data will be analysed as planned in this statistical analysis plan.

All efficacy endpoints will be analysed for the FAS analysis set, and as a sensitivity the analyses for the primary endpoint will be repeated for the PP population.

Categorical data will be summarised using counts and percentages, while continuous data will be presented using the number of patients (N), mean, standard deviation, median, interquartile range, minimum and maximum. All efficacy endpoints will be listed in patient data listings.

9.2 Primary Endpoint

- Vasopressor- and mechanical ventilator-free days (P&VFDs) up to day 30

This composite endpoint is defined as the number of days (reported to one decimal place (0.0 to 30.0)) from start of treatment with the investigational medicinal product (IMP) [selepressin or placebo] to 30.0 days thereafter during which the patient is: 1) alive; 2) free of treatment with intravenous vasopressors; and 3) free of any invasive mechanical ventilation (see definition below).

Patient Death

By definition, any patient that dies within this 30-day period will be assigned zero P&VFDs, even if there is a period during which the patient is alive and free of both vasopressor treatment and mechanical ventilation.

Definition of "Free of Vasopressors"

Free of vasopressors means less than 60 minutes during any contiguous 24-hour period (regardless of calendar day). If a patient requires periods of vasopressors longer than 60 minutes in total during any 24-hour period, the intervening intervals during which they are free of vasopressors will not be included in the period free of vasopressors in the determination of the number of P&VFDs. Thus, the period free of vasopressors begins at the end of the last use of vasopressors that was either: 1) longer than 60 minutes in duration; or 2) part of greater than 60 minutes of use within a contiguous 24-hour period.

Norepinephrine/noradrenaline, phenylephrine, dopamine, epinephrine/adrenaline, vasopressin, terlipressin, and IMP (i.e. selepressin and placebo) all constitute a vasopressor for the purpose of the primary analysis.

Vasopressor use due to anaesthesia or procedure-induced hypotension during and up to three hours after a surgery or procedure (including bedside) is exempt from this rule (i.e. such use of vasopressors would not affect the calculation of P&VFDs).

Definition of "Free of Mechanical Ventilation"

Mechanical ventilation is defined as use of endotracheal or tracheostomy tube assisted ventilation (>5 cm H_2O continuous positive airway pressure and >5 cm H_2O of pressure support from the ventilator in tracheostomy patients). End of mechanical ventilation is defined as: 1) extubation of intubated patients or 2) ≤5 cm H_2O continuous positive airway pressure and ≤5 cm H_2O of pressure support from the ventilator in tracheostomy patients. If non-invasive ventilation by mask or bag (>5 cm H_2O of pressure support) is deployed to avoid (re)intubation, it also counts as mechanical ventilation. However, all other uses of non-invasive ventilation such as chronic night-time use of positive airway pressure for chronic obstructive pulmonary disease (COPD) or sleep apnea does not count as mechanical ventilation (regardless of pressure). Free of mechanical ventilation means less than 60 minutes during any contiguous 24-hour period (regardless of calendar day). If a patient requires mechanical ventilation for periods longer than 60 minutes in total during any 24-hour period, the intervening intervals during which they are not receiving mechanical ventilation will not be included in the period free of mechanical ventilation in the determination of the number of P&VFDs. Thus, the period free of mechanical ventilation begins at the end of the last use of mechanical ventilation that was either: 1) longer than 60 minutes in duration; or 2) part of greater than 60 minutes of use within a contiguous 24-hour period.

The use of mechanical ventilation associated with anaesthesia or procedural sedation during and up to three hours after a surgery or procedure (including bedside) is exempt from this rule (i.e. such use of mechanical ventilation would not affect the calculation of P&VFDs).

It is important to note that the determination of freedom from vasopressors and freedom from mechanical ventilation are made separately; in other words, periods of vasopressor use and mechanical ventilation are not combined when determining whether 60 minutes of use has occurred within a 24-hour period.

Missing data during the time of hospitalization will be imputed using a worst case approach taking into account previous and subsequent starting and stopping times of vasopressor administration and mechanical ventilation (see Figure 2). If only the stop date but not time is given, the imputed time will be midnight of that date (example A), unless a subsequent starting time was recorded prior to midnight in which case the imputed time would be the start time of the subsequent record (example B). If neither stop date nor time is given, the imputed stop time will be the start date and time of the subsequent recording. Likewise, missing start dates and times would be imputed as worst case scenarios, i.e. is the patient found to be on mechanical ventilation with a date but no time for intubation, the imputed start time would be recorded as 00:01 of that day or the stop date of a preceding recording on that same date, whichever occurs last. If both start date and time is missing, the imputed start time would be the date and time of the preceding stop time recorded. In case of data being completely missing from a certain time point and onwards, the "last status carried forward" imputation (example C) will be applied. If a patient was last seen on either ventilator or vasopressors, it is assumed that the patient remained so, and is imputed to a value of 0 (zero) P&VFDs. If the patient was last seen off ventilator and vasopressors, it is also assumed that the patient remained so in the remaining 30-day period. If the patient was last seen (alive) on e.g. day

10 and at that point had been off both ventilator and vasopressors for three days, a value of 23 P&VFDs is imputed.

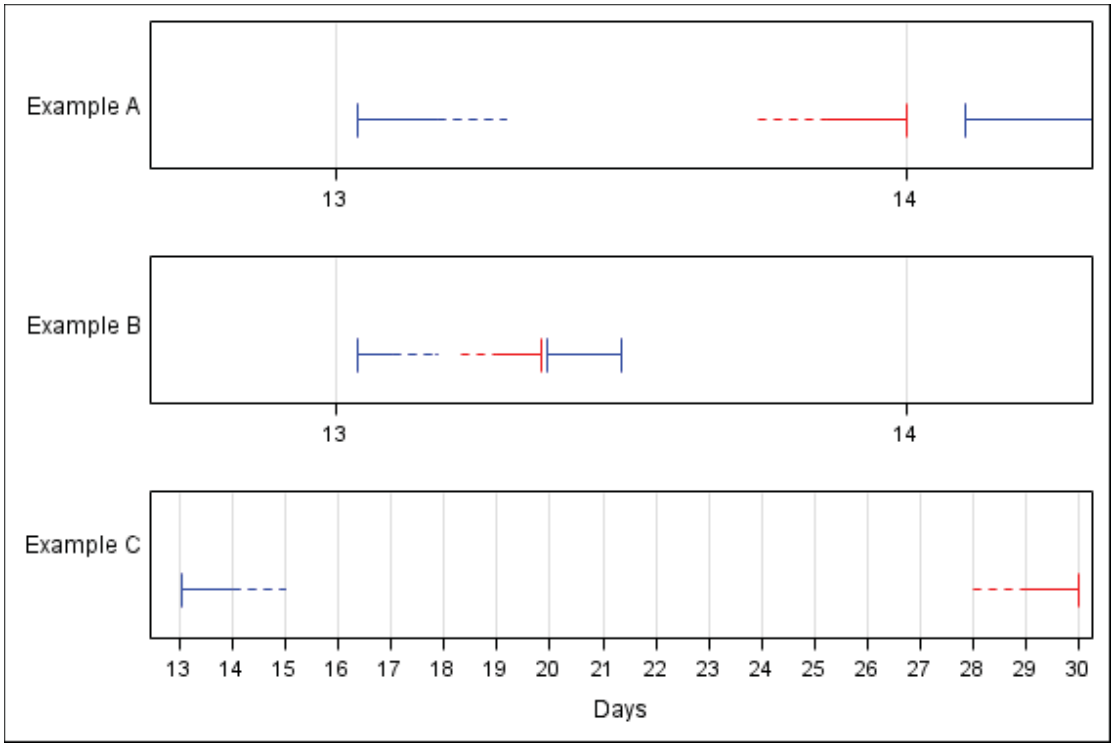

Figure 2: Missing Data Imputation of P&VFDs (Examples)

9.2.1 Primary Variable Analysis

The primary endpoint, P&VFDs, will be analyzed using a van Elteren test, stratified by need for ventilation (Yes/No), time from onset of shock (onset of any vasopressor) to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min).

The primary analysis will compare all patients on all selepressin arms from both parts of the trial (pooled together and treated as a single arm) to all patients on the placebo arm from both parts of the trial (see Appendix 7 for a discussion on treatment estimate bias).

The primary analysis will be a test of superiority using a two-sided 5% significance level test.

Treatment effects will be estimated assuming a negative binomial distribution (to allow for possible overdispersion in a Poisson distribution) for the quantity (30 minus P&VFDs) for survivors, and a binomial distribution to model the probability of surviving. Both models adjusted for need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) (see Appendix 3 for details). For completeness, the proportion of patients dying, and the P&VFDs for survivors will also be presented.

Furthermore, P&VFDs will be tabulated by treatment arm (including pooled active treatment arms), and presented graphically by histograms and cumulative distributions functions.

The success (statistical/clinical significance) of the trial will be based upon the comparison of the analysis above (all patients on all selepressin arms from both parts of the trial (pooled together and treated as a single arm) compared to all patients on the placebo arm from both parts of the trial).

9.2.2 Sensitivity Analyses

The primary analysis will be repeated for the PP analysis set.

As the adaptations of the trial provide a conservative estimate of the p-value, sensitivity p-values will be provided using post-simulation bootstrap calculations (see Appendix 6 for details).

In order to check for consistency the primary endpoint treatment differences will, as a minimum, be estimated and presented by forest plots for the following subgroups

- region (US/Canada vs. Europe)
- age (<65, 65-74, 75-84, ≥85)
- gender
- race/ethnicity

Furthermore, the primary endpoint will be stratified by severity of the patients, with risk of dying as indicator of severity (Figure 3). Mortality (the risk of dying) will be predicted by a logistic regression model, with relevant baseline characteristics as covariates (e.g. the individual SOFA scores and age). The model used to generate the predicted risk (for all patients) will be based on patients in the placebo arm only, as the risk of dying should reflect the severity in the absence of selepressin. Stratified by the risk of dying (intervals of 20% if suitable, based on the mortality rates in the covariate categories in the model), the treatment effect of the primary endpoint will be presented graphically, in order to visually inspect whether the average treatment effect is distributed evenly across the severity of patients.

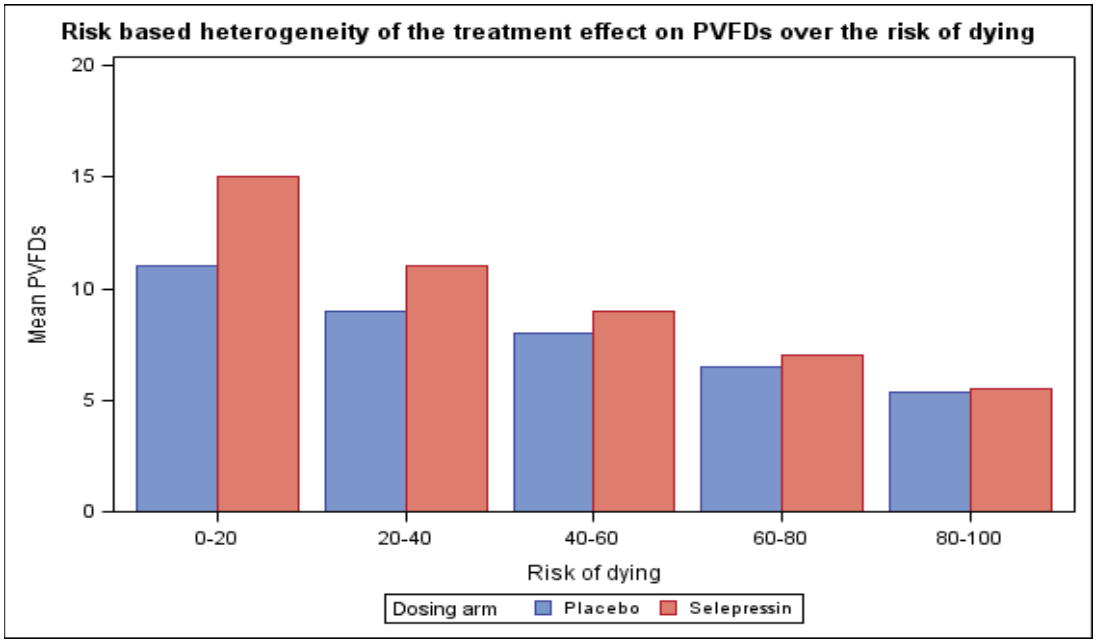

Figure 3: Risk Based Heterogeneity of the Treatment Effect on P&VFDs over the Risk of Dying (an Example)

The impact and robustness of the imputation of missing data will be checked by analysing data in the following ways

- excluding all patients with missing/imputed data
- imputing the 30-day P&VFD status for patients lost to follow up (or otherwise withdrawn from trial) using the observed ratio of P&VFDs at time of lost to follow up (or time of withdrawal), to the same proportion for a 30-day status

For this analysis the 30-day P&VFD status for patients lost to follow up (or otherwise withdrawn from trial) will be imputed so that the 30-day ratio of P&VFDs is equal to the ratio of P&VFDs at time of lost to follow up (or time of withdrawal). E.g. a patient being lost to follow up at Day 15 with 4 P&VFDs (a ratio of 4/15 P&VFDs per days observed) will be imputed to 8 P&VFDs at Day 30 (equivalent ratio 8/30 = 4/15). Patients having zero P&VFDs at time of lost to follow up will be imputed to a value of zero P&VFDs.

- tipping point analysis

The tipping point analysis will compare all possible combinations of 'best case' and 'worst case' scenarios between placebo and selepressin (Figure 4) for patients lost to follow up (or otherwise withdrawn from trial). Best case being an imputation assuming the remaining days off ventilator and vasopressors, and worst case being an imputation of 0 P&VFDs. Let N_p and N_s be the number of patients in the placebo and selepressin arms with missing data. The tipping point analysis will compare all combinations (from 0 to N_p) of X patients on placebo imputed best case and $N_p - X$

imputed worst case, to Y patients on selepressin imputed best case and N_s – Y imputed worst case. I.e all N_p +1 times N_s +1 combinations will be analysed for the primary endpoint. Since the 'best case' is not the same for all patients (depending on when they were last seen off both ventilator and vasopressors) there are multiple outcomes within each combination. For each combination, the average P-value of the multiple outcomes will be plotted in the tipping point analysis. Below is an example of a tipping point analysis of 25 placebo patients vs. 40 selepressin patients with imputed values. The x- and y-axis displays the number of patients with the 'best case' imputed. In the example below the red area displays the non-significant p-values, indicating that one would have to impute almost all placebo patients to a 'best case' and almost all selepressin to a 'worst case' in order to get non-significant p-values, and hence 'proving' the robustness of the imputation method.

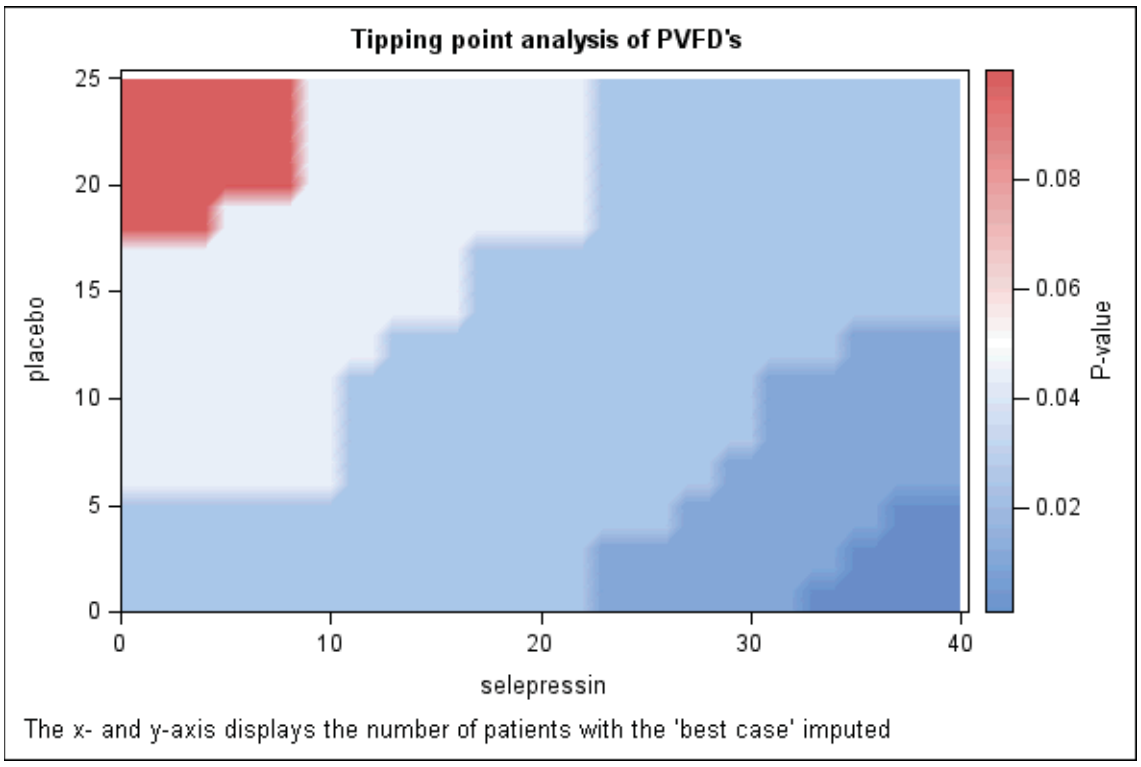

Figure 4: Tipping Point Analysis of P&VFDs (an Example)

Also, to make sure that the use of vasopressor in each group is not simply being replaced by an increased use of inotropic agents (e.g. dobutamine, milrinone, levosimendan, and amrinone), there will be a sensitivity analysis of the primary endpoint in which the use of inotropic agents will count as vasopressor use.

9.2.3 Additional Analyses

The primary analysis will be repeated for:

- the selected arm only, i.e. comparing all patients on the selected arm (from part 1) from both parts of the trial (pooled), to all patients on the placebo arm from both parts of the trial.

- data from part 2 only, i.e. comparing the selected arm to placebo on data from part 2 only

9.3 Secondary Endpoints

For the purpose of a possible label inclusion, the Hochberg procedure (4) for adjustment on multiplicity will be implemented to selected key secondary endpoints. Only if the primary efficacy analysis leads to a statistically significant result at the (one-sided) 2.5% level, then the Hochberg procedure which is described below is applied to selected key secondary analyses. If the primary efficacy analysis does not result in statistical significance at the (one-sided) 2.5% level, then statistical significance (for the purpose of a possible label inclusion only) will not be declared for any of the key secondary analyses, regardless of their p-values.

The selected key secondary endpoints aimed at further demonstrating treatment effect are:

- All-cause mortality (defined as the fraction of patients that have died, regardless of cause, by the end of Day 90)
- Renal replacement therapy-free days up to Day 30 (excluding patients on RRT for chronic renal failure at time of randomisation)
- ICU-free days up to Day 30

In this application of the Hochberg procedure there are three hypothesis tests of superiority for each of the selected secondary endpoints. The target alpha level is (one-sided) 2.5%. The Hochberg procedure is as follows:

- Order the p-values from the smallest to the largest value, $p(1) < p(2) < p(3)$, with corresponding null hypothesis $H_{(1)}$, $H_{(2)}$, and $H_{(3)}$.
- Start with the highest p-value. If $p(3) < 2.5\%$ (one-sided), then stop and declare all three comparisons significant at the 2.5% (one-sided) level (i.e. reject $H_{(1)}$, $H_{(2)}$, and $H_{(3)}$). Otherwise, accept $H_{(3)}$ for the endpoint related to p(3), and go to p(2) the second highest p-value.
- If $p(2) < 2.5/2 = 1.25\%$ (one-sided), then stop and declare significance for $H_{(1)}$ and $H_{(2)}$. Otherwise, accept $H_{(2)}$, for the endpoint related to p(2), and go to p(1) the lowest p-value.
- If $p(1) < 2.5/3 = 0.833\%$ (one-sided), then stop and declare significance for $H_{(1)}$. Otherwise, accept $H_{(1)}$, for the endpoint related to p(1).

Regardless of the statistical significance declared according to the Hochberg procedure, all analysis will be included and presented in the statistical report.

As for the primary analysis, the primary comparison (which determines the success, i.e. statistical and clinical significance) for the secondary efficacy endpoints is between all patients on all selepressin arms from both parts of the trial (pooled together and treated as a single arm) and all patients on the placebo arm from both parts of the trial.

As an additional analysis, all secondary efficacy analyses will, as for the primary, be repeated for:

- the selected arm only, i.e. comparing all patients on the selected arm (from part 1) from both parts of the trial (pooled), to all patients on the placebo arm from both parts of the trial.
- data from part 2 only, i.e. comparing the selected arm to placebo on data from part 2 only

All free-days endpoints will be reported to one decimal place.

9.3.1 Organ Dysfunction

9.3.1.1 Vasopressor-free Days up to Day 30

Vasopressor-free Days up to Day 30 will be defined and analyzed in a similar manner as the primary endpoint, with the van Elteren test stratified by need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 μg/min).

9.3.1.2 Mechanical Ventilator-free Days up to Day 30

Ventilator -free Days up to Day 30 will be defined and analyzed in a similar manner as the primary endpoint, with the van Elteren test stratified by need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 μg/min).

9.3.1.3 Duration of Septic Shock up to Day 30

Duration of septic shock is defined as the cumulative periods (>1 hour) from start of IMP until Day 30, on IMP or vasopressors. Vasopressor use due to anaesthesia / procedure-induced hypotension during - and up to three hours after - surgery / procedure (including bedside) is exempt from this rule

For patients withdrawn (in the survivors analysis) or dying (in the non-survivors analysis) while still in septic shock, the duration will be based on the data available up until the time of withdrawal or death.

Duration of septic shock will be analyzed separately for survivors, non-survivors (within the first 30 days) and overall, comparing treatment arms by a negative binomial model with time from onset of shock to start of treatment and norepinephrine/noradrenaline requirement at baseline (μg/kg/min) as covariates, and treatment and need for ventilation (Yes/No) as factors. The estimated treatment difference (to placebo) with a 95% confidence interval will be presented. As duration of septic shock is derived with two decimal points (with values in the range of 0.00, 0.01, ..., 30.00), duration of septic shock MAY be transformed (for model stability) to integer values (with values in the range of 0, 1, ..., 3000) by multiplying with 100. Treatment estimates will be back transformed

Some patients will get out of shock prior to Day 30 (and stay alive until Day 30), some will get out of shock and die later on (prior to Day 30), others will die or be withdrawn while still in shock (prior to Day 30), and the remaining few will not get out of shock prior to Day 30. This means that if mortality and withdrawal rates vary between treatment arms, the results of the analysis for the overall population will be influenced by the skewed mortality and withdrawal rates. Hence, for the overall population, the distribution of duration of shock (time to out of shock), will be presented graphically as competing risks between 'time to out of shock', 'withdrawn while in shock' and 'dying while in shock'. Further, Kaplan-Meier (sub)-graphs on 'time to death' and 'time to withdrawal' will be presented for those getting out of shock (for which some will die or be withdrawn later on, prior to Day 30). This is done in order to elucidate any skewness in mortality and withdrawal rates, influencing the results of the analysis.

Furthermore, duration of septic shock will be tabulated by treatment arm (including pooled active treatment arms).

In case the model assumptions does not hold, covariates will be dichotomised (in same manner as they have been dichotomised for other endpoint analyses) and duration of septic shock will be analysed using a stratified permutation test (Monte Carlo estimate of p-value). Treatment effects (and treatment difference) will be the raw (unadjusted/non-stratified) means. The confidence intervals for the individual treatment arms will be derived by bootstrapping 5000 samples (within treatment group) with same number of observations (within treatment group), and taking the 2.5 and 97.5% percentiles from the distribution of the 5000 means from the bootstrapped samples. The confidence interval for the treatment difference will be derived via significance testing (based on the stratified permutation test), i.e. the confidence interval will be constructed by including the 95% confidence region for all those values for which the significance test of the hypothesis that the true value is the given value is not rejected at a 5% significance level (e.g. if we test is treatment difference=1 and p-value >0.05 then we include 1 in the confidence interval, and so on). The seed for both the Monte Carlo derived p-value and the bootstrapped confidence intervals will (appropriately) be seed=133.

9.3.1.4 Duration of Mechanical Ventilation up to Day 30

Duration of mechanical ventilation is defined as the cumulative periods (>1 hour) from start of IMP until Day 30, on mechanical ventilation. Mechanical ventilation during - and up to three hours after - surgery / procedure is exempt from this rule.

For patients withdrawn or dying while still on mechanical ventilation, the duration will be based on the data available up until the time of withdrawal or death.

Duration of mechanical ventilation will be analyzed separately for survivors, non-survivors (within the first 30 days) and overall using a zero inflated negative binomial model with time from onset of shock to start of treatment, and norepinephrine/noradrenaline requirement at baseline (µg/kg/min) as covariates, and treatment and need for ventilation as factors. As duration of ventilation is derived with two decimal points (with values in the range of 0.00, 0.01, ..., 90.00), duration of ventilation

will be transformed (for model stability) to integer values (with values in the range of 0, 1, ..., 9000) by multiplying with 100. Treatment estimates will be back transformed.

In case the model assumptions for duration of ventilation does not hold, the permutation test specified in Section 9.3.1.3 will be used including an analysis of the expected proportion of patients with 0 duration, and estimates of the duration for those with a positive duration.

9.3.1.5 Daily Overall and Individual Organ (Cardiovascular, Respiratory, Renal, Hepatic, Coagulation) Scores using the Modified Sequential Organ Failure Assessment (SOFA) Scores Until ICU Discharge

Last observation carried forward (LOCF) will be used for missing SOFA scores on Days 2-7. No LOCF for Day 1 (as previous value is baseline). Patients dying will be imputed with a worst possible outcome, i.e. a value of 4 for each individual SOFA score.

Daily overall (modified) and individual SOFA scores will be compared between treatment arms up until Day 7 using a repeated measures ANCOVA model with baseline SOFA score as covariate, treatment, time and treatment by time interaction as factors, baseline SOFA score by time interaction, and patient as the experimental unit. The estimated treatment difference (to placebo) with a 95% confidence interval will be presented.

Furthermore, daily overall and individual SOFA scores will be tabulated by treatment arm (including pooled active treatment arms).

9.3.1.6 Incidence of New Organ Dysfunctions and New Organ Failures (Based on the SOFA score) up to Days 7 and 30

Incidence of new organ failures is defined as a change in any of the individual SOFA scores from (0,1,2) at baseline to (3,4) post baseline up until the end of the period (Day 7 or 30) (if the SOFA scores goes from (0,1,2) to (3,4) and back to (0,1,2) again within the period, that will still count as a new organ failure). If a patient dies within the period, he is considered to fail on all organs, and the number of new organ failures will be all organs except those already failed at baseline. Patients withdrawn within the period will be evaluated based on the data available at time of withdrawal.

Incidence of new organ dysfunction is defined as an increase ≥ 1 from baseline to post baseline up until the end of the period (e.g. going from 1 to 2) in any of the individual SOFA scores. Patients with an individual SOFA score of 4 at baseline can per default not have a new organ dysfunction. If a patient dies within the period, he is considered to have dysfunction on all organs, and the number of new organ dysfunctions will be all organs except those already having a score of 4 at baseline. Patients withdrawn within the period will be evaluated based on the data available at time of withdrawal.

As the SOFA score is only collected for patients still in the ICU, it is assumed that as soon as the patient leaves the ICU, the patient will not experience any new organ dysfunctions or organ failures. Unless of course the patient dies.

Incidence of at least one new organ failure will be analyzed for any new organ failure (across all organ systems) and by individual organ systems, and compared between treatment arms using a logistic regression model with age, modified SOFA score and norepinephrine/noradrenaline requirement at baseline (µg/kg/min) as covariates and gender and treatment arm as factors, presenting odds ratios with 95% confidence intervals.

Incidence of at least one new organ dysfunction will be analyzed for any new organ dysfunction (across all organ systems) and by individual organ systems, and will be analyzed as above for new organ failures.

The number of new organ failures and the number of new organ dysfunctions will be compared between treatment arms using a negative binomial model with age, modified SOFA score and norepinephrine/noradrenaline requirement at baseline (µg/kg/min) as covariates, and gender and treatment as factors. The estimated treatment difference (to placebo) with a 95% confidence interval will be presented. The 95% confidence interval for the difference in proportions between treatment groups will be constructed using the delta method (see Appendix 4 for details).

Furthermore, incidence of any new organ failure and any new organ dysfunction, and the number of new organ failures and new organ dysfunctions will be tabulated by treatment arm (including pooled active treatment arms).

9.3.1.7 Renal Replacement Therapy (RRT)-free Days up to Day 30 (excluding patients on RRT for chronic renal failure at time of randomisation)

RRT-free Days is defined as for the primary endpoint with free of treatment with any form of renal replacement therapy defined as continuous renal replacement therapy, intermittent hemodialysis or peritoneal dialysis.

RRT-free Days will be analyzed excluding patients on RRT for chronic renal failure at time of randomisation.

RRT-free Days up to Day 30 will be analyzed in a similar manner as the primary endpoint, with the van Elteren test stratified by need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and baseline creatinine (< or ≥150 µmol/L).

9.3.1.8 Incidence of RRT up to Day 30 (counting patients who die as on RRT and excluding patients on RRT for chronic renal failure at time of randomisation)

RRT is defined as any form of renal replacement therapy defined as continuous renal replacement therapy, intermittent hemodialysis or peritoneal dialysis. In order to ensure that any reduction in incidence of RRT is not caused by an increase in mortality, all patients dying within the 30-day period will be counted as on RRT. For patients withdrawn prior to Day 30, incidence of RRT will be based on the data available up until the time of withdrawal.

Incidence of RRT will be analysed by a logistic regression model with time from onset of shock to start of treatment, baseline creatinine and norepinephrine/noradrenaline requirement at baseline (µg/kg/min) as covariates, and treatment and need for ventilation as factors. The 95% confidence interval for the difference in proportions between treatment groups will be constructed using the delta method (see Appendix 4 for details). Patients already on RRT for chronic renal failure at time of inclusion will be excluded from the analysis of incidence of RRT.

Non-inferiority will be claimed if the upper limit of the two-sided 95% CI of the adjusted difference in proportions is less than 20% of the estimated incidence of RRT in the placebo group. Superiority can be claimed (5) if the upper limit is less than 0.

I.e., let $\widehat{p_S}$ and $\widehat{p_P}$ be the estimated incidences of RRT in the combined selepressin groups and the placebo group respectively. Non-inferiority will then be claimed if

$$\widehat{p_S} - \widehat{p_P} + 1.96 * \sqrt{var(\widehat{p_S} - \widehat{p_P})} < 0.2 * \widehat{p_P}$$

and superiority will be claimed if

$$\widehat{p_S} - \widehat{p_P} + 1.96 * \sqrt{var(\widehat{p_S} - \widehat{p_P})} < 0$$

A forest plot including the non-inferiority tests for all analyses will be presented.

Furthermore, incidence of RRT will be tabulated by treatment arm (including pooled active treatment arms).

A subgroup analysis will be performed on patients without acute RRT at baseline.

9.3.1.9 Duration of RRT up to Day 90 (excluding patients on RRT for chronic renal failure at time of randomisation)

Duration of RRT is defined as the cumulative periods with RRT (continuous renal replacement therapy, intermittent hemodialysis or peritoneal dialysis) and will be analyzed excluding patients on RRT for chronic renal failure at time of randomisation.

For patients withdrawn or dying while still on RRT, the duration will be based on the data available up until the time of withdrawal or death.

Duration of RRT will be analyzed as for duration of ventilation (also by transforming to integers) in Section 9.3.1.4 with time from onset of shock to start of treatment, baseline creatinine and norepinephrine/noradrenaline requirement at baseline (µg/kg/min) as covariates, and treatment and need for ventilation as factors. The supportive competing risk and Kaplan-Meier graphs described in Section 9.3.1.4 will not be performed as it makes no sense since RRT durations are short and repetitive. It is more meaningful to just look at 90 day mortality and withdrawal rates to take eventual skewed mortality and withdrawal rates into account.

9.3.2 Morbidity/Mortality

9.3.2.1 Intensive Care Unit (ICU)-free Days up to Day 30

ICU -free Days up to Day 30 will be defined and analyzed in a similar manner as the primary endpoint, with the van Elteren test stratified by need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and baseline creatinine (< or ≥150 µmol/L).

A sensitivity analysis will be performed where the definition of time spent in the ICU will include first admission to emergency department, as some patients will be enrolled and have first IMP treatment in the emergency department.

9.3.2.2 ICU Length of Stay up to Day 30

ICU length of stay is defined as the cumulative periods spent in ICU from start of IMP to 30 days after.

For patients withdrawn or dying while still in ICU up to Day 30, the duration will be based on the data available up until the time of withdrawal or death.

ICU length of stay will be analyzed between treatment groups using the permutation test specified in Section 9.3.1.3 stratified by the combination of time from onset of shock to start of treatment (< or ≥6 hours), baseline creatinine (< or ≥150 µmol/L), norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and need for ventilation (Yes/No).

A sensitivity analysis will be performed where the definition of time spent in the ICU will include first admission to emergency department, as some patients will be enrolled and have first IMP treatment in the emergency department.

9.3.2.3 All-cause Mortality (Defined as the Fraction of Patients That Have Died, Regardless of Cause, by the end of Day 30, Day 90, and Day 180)

Mortality will be analysed by a logistic regression model with the individual SOFA scores and age as covariates and treatment arm as factor. The 95% confidence interval for the difference in proportions between treatment groups will be constructed using the delta method as for incidence of RRT (section 9.3.1.8). There will be no imputations for mortality.

Non-inferiority will be claimed if the upper limit of the two-sided 95% CI of the adjusted difference in proportions is less than 30% of the estimated incidence of mortality in the placebo group. Superiority can be claimed (5) if the upper limit is less than 0.

I.e., let $\widehat{p_S}$ and $\widehat{p_P}$ be the estimated incidences of mortality in the combined selepressin groups and the placebo group respectively. Non-inferiority will then be claimed if

$$\widehat{p_S} - \widehat{p_P} + 1.96 * \sqrt{var(\widehat{p_S} - \widehat{p_P})} < 0.3 * \widehat{p_P}$$

and superiority will be claimed if

$$\widehat{p_S} - \widehat{p_P} + 1.96 * \sqrt{var(\widehat{p_S} - \widehat{p_P})} < 0$$

Assuming observed mortality rates of 20-25% in the placebo group, a non-inferiority limit of 30% corresponds to a maximum observed mortality rate of 2-3% in the combined selepressin groups in order for selepressin to be non-inferior to placebo. A forest plot including the non-inferiority tests for all analyses will be presented.

Furthermore, mortality will be tabulated by treatment arm (including pooled active treatment arms), and the time to death presented graphically by a Kaplan-Meier plot.

9.3.3 Health-Related Quality of Life

9.3.3.1 Change in utility, based on EQ-5D-5L, up to Day 180

EQ-5D-5L will be analyzed by the index value, the overall QALY (Quality-Adjusted Life Years) at Day 30 and 180 (see Appendix 1 for details), and the VAS score.

The QALY scores will NOT be adjusted to e.g. a half yearly time scale at Day 180.

As the QALY is not defined for patients with all remaining values missing, and hence also not defined for those dead, the analyses will automatically only be analyzed for those surviving up until Day 30 and 180 respectively.

For patients with missing baseline index value, the QALY score will also be set to missing. For robustness, a sensitivity analyses will be performed, imputing the missing baseline scores with the overall mean of the baseline health index. Baseline is the timing prior to acute admission.

The QALY at Day 30 and 180 will be compared between treatment arms using an ANCOVA model with baseline health index as covariate, and treatment as factor. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented.

The index value and VAS scores will be analysed separately for survivors and non-survivors at Day 180 (since all non-survivors will have non-random missing values, and hence would artificially inflate the mean estimates if survivors and non-survivors were analysed together) and will be compared between treatment arms using a repeated measures ANCOVA model with baseline health index/VAS score as covariate, treatment, time and treatment by time interaction as factors, baseline index/VAS score by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented for Day 30, 60, 90 and 180. There will be no imputations for missing values.

Furthermore, the QALY, index value and VAS scores will be tabulated treatment arm (including pooled active treatment arms), and the index value and VAS scores will be presented graphically.

9.3.4 Fluid Balance

9.3.4.1 Daily and Cumulative Fluid Balance Until ICU Discharge (for a Maximum of 7 Days)

Fluid overload is defined as fluid balance volume (L) as a percentage of baseline weight. E.g. if a patient weighs 90 kg at baseline and has a fluid balance of 9 liters, fluid overload is then 100% * 9L / 90kg = 10%.

Fluid balance (as a rate of time) and cumulative fluid balance (total volume) will be presented both unadjusted and adjusted for weight.

Baseline fluid balance (as a rate of time) and baseline fluid balance volume (mL) will be based on the time from onset of sepsis induced hypotension to start of IMP.

All analyses will be presented for ‘all patients’ and for ‘patients in ICU throughout Day 0-7’.

Fluid balance, cumulative fluid balance, fluid overload and cumulative fluid overload will all be compared between treatment arms using a repeated measures ANCOVA model with baseline (baseline fluid balance volume (mL) or baseline fluid overload) as covariate, treatment, time and treatment by time interaction as factors, baseline score by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented.

Fluid overload and balance will be based on a sectioning of the intervals between the actual sampling time points on baseline, Day 1, 2, etc., in order to be able to make 24 hour intervals from start of IMP.. For Days 1, 2, etc. sampling time points will be every 24 hours from start of IMP. Patients withdrawn or dead will be set to missing.

Endpoints (absolute values and change from baseline) will be tabulated by treatment arm (including pooled active treatment arms), and presented graphically.

Furthermore, in order to assess all patients at Day 3 and 7 (regardless of mortality), the composite endpoint ‘critical edema-free survival’ will be derived at both Day 3 and 7.
A patient will be categorised as being critically edema-free and alive at Day 3 (or 7 respectively) if:

- fluid overload < 10% at Day 3, or
- patient is discharged from ICU and emergency department prior to Day 3 with a fluid overload < 10% at day of discharge (or previous day if discharged before fluid balance measurement), and patient is alive at Day 3.

The patient will be categorised as not being critically edema free or dead if:

- fluid overload ≥ 10% at Day 3, or

- patient is discharged from ICU and emergency department prior to Day 3 with a fluid overload ≥ 10% at day of discharge (or previous day if discharged before fluid balance measurement), or
- patient is dead on Day 3 (prior to fluid balance measurement)

Patients withdrawn at the given time point will be set to missing (less than 20 patients total in the trial are expected to be withdrawn prior to Day 7). The endpoints will be analysed using a logistic regression model with baseline fluid overload as a covariate and treatment arm as a factor, presenting odds ratios with 95% confidence intervals.

9.3.4.2 Daily and Cumulative Urinary Output Until ICU Discharge (for a Maximum of 7 Days)

Urinary output and cumulative urinary output (absolute values) will all be will be compared between treatment arms using a repeated measures ANCOVA model with baseline urinary output volume (mL) as covariate, treatment, time and treatment by time interaction as factors, baseline score by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented.

Urinary output and cumulative urinary output will be presented both unadjusted and adjusted for weight.

All analyses will be presented for ‘all patients’ and for ‘patients in ICU throughout Day 0-7’.

Urinary output will be derived in the same manner as fluid balance.

Endpoints (absolute values and change from baseline) will be tabulated by treatment arm (including pooled active treatment arms), and presented graphically.

9.4 Other Endpoints/Assessments

9.4.1 Hospital-free Days up to Day 90

Hospital stay is defined (from the eCRF) as “Other acute care hospital” or “Still in trial hospital”

Hospital-free Days up to Day 90 will be defined and analyzed in a similar manner as the primary endpoint, with the van Elteren test stratified by need for ventilation (Yes/No), time from onset of shock to start of treatment (< or ≥6 hours), and norepinephrine/noradrenaline requirement at baseline (< or ≥30 μg/min) and baseline creatinine (< or ≥150 μmol/L).

9.4.2 Hospital Length of Stay up to Day 90

Hospital length of stay is defined as the cumulative periods spent in hospital (“Other acute care hospital” or “Still in trial hospital”) from start of IMP to 90 days after.

For patients withdrawn or dying while still in hospital up to Day 90, the duration will be based on the data available up until the time of withdrawal or death.

Hospital length of stay will be analyzed between treatment groups using the permutation test specified in Section 9.3.1.3 stratified by the combination of time from onset of shock to start of treatment (< or ≥6 hours), baseline creatinine (< or ≥150 µmol/L), norepinephrine/noradrenaline requirement at baseline (< or ≥30 µg/min) and need for ventilation (Yes/No).

9.4.3 Patient Residence at Day 30, Day 60, Day 90, and Day 180

Patient location at Day 30, 60, 90 and 180 will be summarized by treatment arm (including pooled active treatment arms). Baseline is the timing prior to acute admission.

Shift tables will be presented at each time point to assess whether patients have returned to their location at enrollment. There will be no imputations of missing values.

9.4.4 Mean Arterial Pressure, Until ICU Discharge (for a Maximum of 7 Days)

Mean arterial pressure will be tabulated by treatment arm (including pooled active treatment arms) and presented graphically (at pre-specified time points, i.e. not when NE/NA infusion changes, and only until ICU discharge/day 7) for both MAP alone, and the difference from MAP to target MAP (which is 65, unless the investigator judges it to be otherwise). There will be no imputations of missing values.

MAP will be displayed both for MAP measurements taken while patients are in and out of septic shock, i.e. while patients are on and off IMP/vasopressors (as defined for the primary endpoint).

Difference to target MAP will only be displayed for MAP measurements taken while patients are in septic shock.

9.4.5 Norepinephrine/Noradrenaline and Other Vasopressor Doses

The cumulative dose of norepinephrine/noradrenaline administered (adjusted for baseline weight) will be derived as an area under the curve (AUC) of the actual norepinephrine doses and actual time points (although analysed at planned time points). Linear interpolation will be used to derive the AUC (taking into account that only patients enrolled before noon will have a measurement at 36 hours after IMP start). Patients dead or withdrawn will be set to missing. The cumulative doses (µg/kg) will be compared between treatment arms using a repeated measures ANCOVA model with baseline dose of norepinephrine/noradrenaline (µg/kg/min) as covariate, treatment, time and treatment by time interaction as factors, baseline dose by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented.

If a patient has missing values and the patient is still in the trial (not dead or withdrawn) it will be assumed that the specific vasopressor was not given and a value of zero will be imputed, unless there is an interval in the timing log covering the exact time point (8 AM and 8 PM is the assumed time point for missing morning and evening collection time points). In that case LOCF will be used, but only within the time interval.

The mean dose and cumulative dose administered will be tabulated by treatment arm (including pooled active treatment arms), and presented graphically.

To assess the time needed to wean the patients off norepinephrine, the percentage change from baseline in norepinephrine dose (µg/kg/min) will be compared between treatment arms using a repeated measures ANCOVA model with baseline dose of norepinephrine/noradrenaline (µg/kg/min) as covariate, treatment, time and treatment by time interaction as factors, baseline dose by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented at 1, 3, 6, 12 and 24 hours after start of IMP treatment.

The same analyses (adjusted for baseline dose of norepinephrine/noradrenaline (µg/kg/min)) will be performed for the following endpoints:

- Catecholamines (defined as the sum of doses of norepinephrine/noradrenaline, epinephrine/adrenaline, dopamine, and phenylephrine) (ug/kg)

- Catecholamines excluding norepinephrine/noradrenaline (ug/kg)

- Vasopressin (U/kg)

For the sum of catecholamine doses we define 100 µg dopamine, 1 µg epinephrine, and 2.2 µg phenylephrine all equivalent to 1 µg norepinephrine.

Also, the number of patients receiving terlipressin will be summarised.

9.4.6 Pulmonary Function (PaO2/FiO2) (in a subset of patients)

Baseline for PaO2/FiO2 will be the values obtained at the last assessment prior to the first dose of IMP. There will be no imputations of missing values.

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point by treatment arm (including pooled active treatment arms).

9.4.7 Arterial Blood Gases and Acid/Base Status (PaO2, PaCO2, pH, SaO2, Bicarbonate, Base Excess), Lactate and Oxygen Saturation in Vena Cava Superior (ScvO2)

Baseline for arterial blood gases, lactate and ScvO2 will be the values obtained at the last assessment prior to the first dose of IMP. There will be no imputations of missing values.

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point by treatment arm (including pooled active treatment arms).

9.4.8 Cytokines, ANG-1 and ANG-2 (in a Subset of 100-350 Patients)

Baseline values will be the values obtained at the last assessment prior to the first dose of IMP. There will be no imputations of missing values.

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point by treatment arm (including pooled active treatment arms).

These analyses will not be repeated for the selected arm only and data from part 2 only.

9.4.9 EVLW and PPI (in a Subset of 100-350 Patients)

Baseline for EVLW and PPI will be the values obtained at the last assessment prior to the first dose of IMP. There will be no imputations of missing values.

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point by treatment arm (including pooled active treatment arms).

These analyses will not be repeated for the selected arm only and data from part 2 only.

9.4.10 Cardiac Output (in a Subset of 100-350 Patients)

Baseline for cardiac output values will be the values obtained at the last assessment prior to the first dose of IMP. There will be no imputations of missing values.

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point by treatment arm (including pooled active treatment arms).

These analyses will not be repeated for the selected arm only and data from part 2 only.

9.4.11 Creatinine Clearance

Creatinine clearance is determined by estimated glomular filtration rate (using creatinine, age, and gender as per Cockcroft-Gault).

Cockcroft-Gault equation: creatinine clearance (mL/min) = ((140 - age in years) x weight (kg)) / creatinine (µmol/L) for women. For men, multiply result by 1.2.

Creatinine clearance will be analyzed up until day 3 as for fluid balance in Section 9.3.4.1, with baseline creatinine clearance as a covariate, treatment, time and treatment by time interaction as factors, and patient as the experimental unit.

10 Safety

10.1 General Considerations

Safety parameters will be evaluated for the safety analysis data set.

All safety summaries will be tabulated by treatment arm (including pooled active treatment arms).

10.2 Adverse Events

Adverse events (AEs) are classified according to the Medical Dictionary for Regulatory Activities (MedDRA). The version of MedDRA will be documented in the clinical report.

Written narratives will be issued for all serious AEs (including deaths) and AEs leading to withdrawal.

A pre-treatment adverse event is any untoward medical occurrence arising or observed between informed consent and administration of the IMP.

A treatment emergent adverse event is any adverse event occurring after the administration of the IMP and within the time of residual drug effect, or a pre-treatment adverse event or pre-existing medical condition that worsens in intensity after start of IMP and within the time of residual drug effect.

The time of residual drug effect is the estimated period of time after the end of the administration of the IMP, where the effect of the product is still considered to be present based on pharmacokinetic, pharmacodynamic or other substance characteristics. A generally accepted time for residual drug effect is 5 half-lives. The terminal half-life of the IMP is expected to be not more than 1.8 hours, and treatment-emergent AEs are defined as AEs occurring after the start of study drug infusion to within 12 hours after study drug infusion is stopped.

A post-treatment adverse event is any adverse event occurring after the residual drug effect period. Missing values will be treated as missing, except for causality, intensity, seriousness, and outcome of adverse events. A “worst case” approach will be used: if causality is missing, the adverse event will be regarded as related to the IMP; if the intensity of an adverse event is missing, the adverse event will be regarded as severe; if seriousness is missing the adverse event will be regarded as serious; if start date is missing or incomplete, worst case will be assumed and the AE will be regarded as treatment emergent (only if the incomplete start date is not compromised). If start date is completely missing, start date will be set as same day as start of treatment. If start date is incomplete, the date closest to start of treatment will be assumed, without compromising the incomplete data available on the start date; if outcome is missing and no date of outcome is present the outcome is regarded as ‘not recovered’.

10.2.1 Overview of Adverse Events

AE overview summary tables will be prepared for treatment-emergent AEs and all AEs (treatment-emergent and non treatment-emergent) during the treatment period, including the number of

patients reporting an AE, the percentage of patients (%) with an AE, and the number of events (E) reported, for the following categories:

- Adverse events
- Deaths
- Serious adverse events
- Adverse events leading to discontinuation of IMP
- Severe and life threatening adverse events
- Adverse drug reactions

10.2.2 Incidence of Adverse Events

Adverse events will be summarised in a Table by SOC and PT for MedDRA. The Table will display the total number of patients reporting an AE, the percentage of patients (%) with an AE, and the number of events (E) reported. AEs will be presented by system organ class (SOC) sorted alphabetically and preferred term (PT) sorted in decreasing frequency of occurrence.

For both treatment-emergent AEs and all AEs (treatment-emergent and non treatment-emergent) during the treatment period, summary tables will be prepared for:

- All adverse events
- Adverse events with an incidence ≥ 5% of patients in any treatment arm
- Non-serious adverse events with an incidence ≥ 5% of patients in any treatment arm
- Critical adverse events (see Appendix 2 for details)
- Adverse events by causality (related/unrelated)
- Adverse events leading to death
- Adverse events by intensity
- Serious adverse events
- Adverse events leading to discontinuation of IMP (related/unrelated)

Supporting data listings will be provided for:

- All adverse events sorted by centre and patient no.
- All adverse events sorted by MedDRA Preferred Term
- Serious adverse events
- Adverse events leading to death
- Adverse events leading to discontinuation of IMP (related/unrelated)
- Post treatment-emergent adverse events.

Furthermore, for both treatment-emergent AEs and all AEs (treatment-emergent and non treatment-emergent) during the treatment period, graphs showing the most frequent adverse events and the difference and 95% confidence interval between placebo and selepressin pooled will be presented for:

- All adverse events
- Adverse events with an incidence ≥ 5% of patients in any treatment arm
- Critical adverse events (see Appendix 2 for details)
- Adverse events leading to death
- Serious adverse events

10.3 Safety Laboratory Variables

Baseline for all laboratory analyses will be the values obtained at the last assessment prior to (or at) the first dose of the investigational medicinal product (IMP). End of treatment period will include the last post-baseline observation during the trial up until Day 30.

Laboratory variables will be grouped under “Haematology”, “Clinical Chemistry” or “Coagulation”.

10.3.1 Summary Statistics

Mean change and mean percentage (%) change from baseline at end of treatment period will be presented for each laboratory variable. In addition, descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point for each laboratory variable.

Furthermore, summary tables will be presented, displaying by time and laboratory parameter, the number of patients with a clinically significant result assessed as unanticipated.

Also, a summary table will be prepared for selected laboratory variables that display the number and percentage of patients in each treatment arm with X% increments (increase or decrease) from baseline at pre-specified timepoints and end of treatment period. The following categories for summary tables are defined:

- ≤ -3*X%: Values with more than 3*X% decrease from baseline
- > -3*X% - -2*X%: Values with 2-3*X% decrease from baseline
- > -2*X% - -1*X%: Values with 1-2*X% decrease from baseline
- > -1*X% - 0*X%: Values with 0-1*X% decrease from baseline
- < 0*X%- < 1*X%: Values with 0-1*X% increase from baseline

- 1*X%- < 2*X%: Values with 1-2*X% increase from baseline
- 2*X%- < 3*X%: Values with 2-3*X% increase from baseline
- ≥ 3*X%: Values with more than 3*X% increase from baseline

The following key laboratory variables will be summarised:

- Hemoglobin: 20% change
- WBC: 50% change
- Platelets: 50% change
- Creatinine: 20% change
- Sodium: 5% change
- Lactate: 50% change

10.3.2 Data Listings

Data listings will be prepared by centre, treatment arm, patient and time-point (including baseline) displaying all laboratory values for all patients.

10.4 Vital Signs (including CVP)

Baseline for all vital signs analyses will be the values obtained at the last assessment prior to the first dose of IMP.

10.4.1 Summary Statistics

Descriptive statistics, i.e., the number of patients with data, mean (standard deviation), median, interquartile range, minimum, and maximum values, will be presented for observed values and change from baseline at each time-point for each vital signs variable.

Furthermore, summary tables will be presented, displaying by time and vital signs parameter, the number of patients with a clinically significant result assessed as unanticipated.

Also, a summary table will be prepared for each vital signs variable that display the number and percentage of patients in each treatment arm with X% increments (increase or decrease) from baseline. The following categories for summary tables are defined:

- ≤ -3*X%: Values with more than 3*X% decrease from baseline
- > -3*X% - -2*X%: Values with 2-3*X% decrease from baseline
- > -2*X% - -1*X%: Values with 1-2*X% decrease from baseline
- > -1*X% - 0*X%: Values with 0-1*X% decrease from baseline
- < 0*X%- < 1*X%: Values with 0-1*X% increase from baseline

- 1*X%- < 2*X%: Values with 1-2*X% increase from baseline
- 2*X%- < 3*X%: Values with 2-3*X% increase from baseline
- ≥ 3*X%: Values with more than 3*X% increase from baseline

The following % changes will be summarised:

- Heart rate: 50% change
- Blood pressure: 25% change

10.4.2 Data Listings

Data listings will be prepared by centre, treatment arm, patients and time-point (including baseline) displaying all vital signs values for all patients with an indication of abnormal values.

10.5 Episodes of Hypotension

Descriptive statistics of number of patients with episodes of hypotension and the total length of periods with hypotension will be summarized by treatment arm.

The total length of periods with hypotension will be summarized for both all patients, and patients having one or more episodes of hypotension.

11 Interim analyses

There will be no interim analyses with the potential to stop the trial early for treatment efficacy. However, once the "burn-in" period in Part 1 (first 200 treated patients) is completed, regular interim analyses will be conducted to improve the efficiency of dose selection and to allow early termination of the part or the trial for futility or for successful dose selection. The following steps will be considered at each interim analysis:

- When 200 patients are treated, the allocation probabilities for the active treatment arms are changed using response-adaptive randomisation (with placebo still 1/3). For the two-thirds of patients assigned to the active arms, the probability that a given active arm is assigned to a patient is proportional to the probability that that arm is the arm with the largest expected number of P&VFD
- Potentially stopping the trial for futility during Part 1. This occurs if no active arm has better than a 5% predictive probability of a significant result in Part 2 if it were to start immediately. This decision can occur at any interim during Part 1 after 200 evaluable patients.
- Potentially ending Part 1 and selecting an active treatment arm to continue to Part 2. This decision can occur at any interim analysis between 300 evaluable and 800 treated patients, and it occurs if some arm has a predictive probability of a successful trial of at least 90% before 800 treated patients, and the threshold drops to 25% for the final Part 1 interim analysis at 800 treated patients. The selected arm is the arm with the largest posterior predictive probability of trial success. This will generally be the best-performing active arm, but if multiple arms are performing equally well, it will be the arm with the lowest dosing level. If Part 1 ends after N patients, then Part 2 will consist of up to 1800 - N evaluable patients.
- If the trial is not stopped for futility or proceeding to Part 2 and active treatment Arm 4 has not yet been approved for assignment of patients, the decision can be made to open up Arm 4. Arm 4 is only opened between 200 evaluable and 600 treated patients and if there is at least a 50% probability that Arm 3 has a higher expected P&VFD than Arm 2 and if data from the lower dosing levels do not suggest any significant safety signals.
- If Part 1 reaches its maximum of 800 treated patients and no arm has a predictive probability of Part 2 success of more than 25%, the trial stops with an inconclusive result and will be interpreted as a standalone Phase 2b trial.

During Part 2, interim analyses will be conducted regularly (until 1600 patients have been treated) to allow early termination of the trial for futility. This occurs if the predictive probability of an overall significant result is less than 5%. In addition, if the predictive probability of observing a more than 2% higher mortality in the active arms compared to placebo is greater than 90% then the trial will stop for futility.

12 Deviations from protocol analysis

See change log.

13 References

1 The EuroQol Group (1990). EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 16(3): 199-208

2 EuroQoL Home page: http://www.euroqol.org/home.html

3 Mehta, C.R. and Pocock, S.J. (2011) Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine*, 30, 3267-3284.

4 Hochberg (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75(4):800-802

5 EMEA. (2000). Points to consider on switching between superiority and non-inferiority. CPMP/EWP/482/99.

14 Tables, Listings and Figures

The document with tables, figures and listings (TLF) shells will be presented in a separate document.

Appendix 1 EQ-5D-5L Quality Adjusted Life Year (QALY)

We calculate Quality Adjusted Life Year (QALY) (1), (2) at Day 30 and 180 in three steps:

(1) A unique EQ-5D-5L health state is defined by combining 1 level from each of the 5 dimensions of EQ-5D-5L. Each health state is referred to in terms of a 5 digit code. For example, state 12345 indicates no problems with mobility, slight problems with washing or dressing, moderate problems with doing usual activities, severe pain or discomfort and extreme anxiety or depression.

(2) Convert each EQ-5D-5L health state into a single EQ-5D-5L index value. The index values are country specific and we will use the value sets for US and apply these values to all patients in this trial.

(3) QALY for a patient is then defined to be the area under the curve (AUC) for a Time (with unit of Year) versus index values. AUC will be calculated by the linear trapezoidal method. See below for a schematic presentation where the y-axis is the index value with y_0, y_1 and y_2, etc. represent the index values at baseline, Day 30, 60, 90 and 180. The x-axis is Time (in Years) and the t_0 is the start of treatment period, i.e. baseline, and t_1 and t_2 are time of the actual Day 30 and 60, respectively, and so on.

If the index value at baseline is missing then we set QALYs at Day 30 and 180 to missing. No LOCF imputation will be used. However, linear interpolation will be used between data points with missing data in between (e.g. t_0 to t_2, if t_1 is missing).

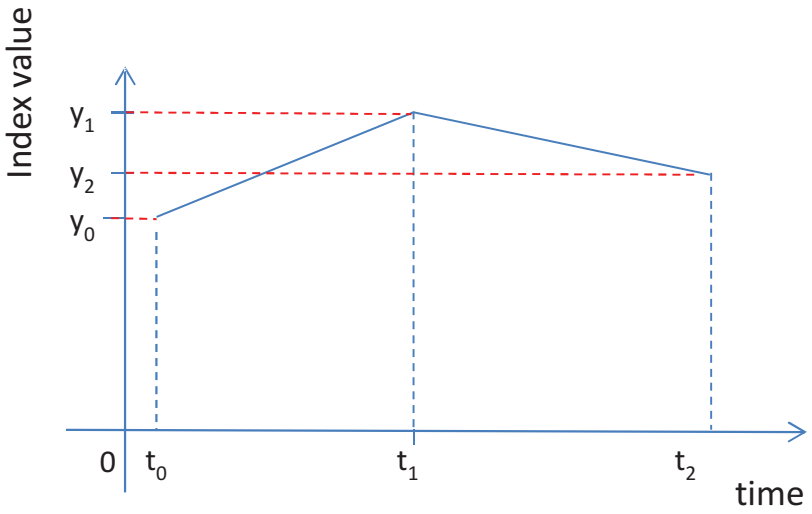

Figure 5: EQ-5D-5L QALY Calculation

Appendix 2 Critical Adverse Events and Classification of Adverse Events Based on Changes in Safety Laboratory Variables

Table 1: Critical Adverse Events

Critical adverse event	MedDRA terms or SMQ used for search
Potential hypersensitivity reactions	Anaphylactic reaction (SMQ) (Narrow Scope) Angioedema (SMQ) (Narrow Scope)
Myocardial infarction / ischemia	Ischaemic heart disease (SMQ) (Broad scope)
Episodes of atrial fibrilation and other cardiac arrythmias	Cardiac Arrhythmias (SMQ) (Broad scope)
Cerebrovascular accident	Haemorrhagic central nervous system vascular conditions (SMQ) (Broad Scope) Ischaemic central nervous system vascular conditions (SMQ) (Broad Scope) Conditions associated with central nervous system haemorrhages and cerebrovascular accidents (SMQ) (Broad scope)
Digital ischemia	Peripheral ischaemia Pallor Peripheral coldness
Renal failure	Acute Renal failure (SMQ)
Mesenteric ischemia	Ischaemic colitis (SMQ) (Narrow scope)
Hepato-biliary adverse events	Drug related hepatic disorders - severe events only (SMQ) (Broad Scope)

Appendix 3 Estimation of Treatment Effects for P&VFDs

Let Y be the number of P&VFDs for both survivors and non-survivors, let p be the probability of surviving, and say that for survivors (30-Y) has a negative binomial distribution (with for Y=0 consuming all values ≥30).

$$P(Y=y)=\begin{cases}(1-p)+p*\sum_{k\geq 30}\frac{\Gamma(d^{-1}+\mathrm{k})}{\Gamma(d^{-1})k!}\left(\frac{\mu*d}{1+\mu*d}\right)^{k}\left(\frac{1}{1+\mu*d}\right)^{d^{-1}} & for\ y=0\\ p*\frac{\Gamma(d^{-1}+30-\mathrm{y})}{\Gamma(d^{-1})(30-y)!}\left(\frac{\mu*d}{1+\mu*d}\right)^{(30-y)}\left(\frac{1}{1+\mu*d}\right)^{d^{-1}} & for\ y>0\end{cases}$$

In the trial it is expected we have a mean of around 5-8 days on vasopressors and mechanical ventilation for those surviving.

Therefore, the probability of getting zero P&VFDs for those surviving

$\sum_{k\geq 30}\frac{\Gamma(d^{-1}+\mathrm{k})}{\Gamma(d^{-1})k!}\left(\frac{\mu*d}{1+\mu*d}\right)^{k}\left(\frac{1}{1+\mu*d}\right)^{d^{-1}}$ is negligible (and can be omitted from the model),

For d=1/16

μ	5	6	7	8	9	10	11
Var =μ+ μ²d	6.56	8.25	10.06	12	14.06	16.25	18.56
P(K≥30)	$1.4*10^{-9}$	$4.2*10^{-8}$	$6*10^{-7}$	$5*10^{-6}$	$2.8*10^{-5}$	$1.2*10^{-4}$	$4*10^{-4}$

μ	12	13	14	15
Var =μ+ μ²d	21	23.56	26.25	26.09
P(K≥30)	$1.1*10^{-3}$	$2.5*10^{-3}$	$5.4*10^{-3}$	$1*10^{-2}$

For $Y_1,...,Y_N$ we then get two likelihood functions,

$$L_1=\prod_{y_i=0}(1-p)$$

and,

$$L_2=\prod_{y_i>0}p*\frac{\Gamma(d^{-1}+30-\mathrm{y})}{\Gamma(d^{-1})(30-y)!}\left(\frac{\mu*d}{1+\mu*d}\right)^{(30-y)}\left(\frac{1}{1+\mu*d}\right)^{d^{-1}}$$

Assuming that p is modeled as $\frac{e^{\beta_1}}{1+e^{\beta_1}}$, and μ as e^{β_2}, the two log-likelihoods then become,

$$l_1=\sum_{y_i=0}\ln\left(\frac{1}{1+e^{\beta_1}}\right)=\sum_{y_i=0}-\ln(1+e^{\beta_1})$$

and,

$$l_2 = \sum_{y_i>0} \beta_1 - \ln(1 + e^{\beta_1}) + \ln\big(\Gamma(d^{-1} + 30 - y_i)\big) - \ln\big(\Gamma(d^{-1})\big) - \ln\big((30 - y_i)!\big) + (30 - y_i)\big(\beta_2 + \ln\big(1 + e^{\beta_2}\big)\big) - d^{-1} * \ln\big(1 + e^{\beta_2}\big)$$

Maximizing the full log-likelihood is thus maximizing $l_1 + l_2$.

And as can be clearly seen, the maximization of l_1 and l_2 with regards to β_1 does not depend on β_2 and vice-versa. Hence, β_1 and β_2 can be estimated separately from two independent models, i.e. a logistic regression for the probability of surviving, and a poisson regression (or negative binomial to allow for overdispersion) for the distribution of days on vasopressors and mechanical ventilation for those surviving.

In practice, and with the model adjusted for A, B and C, this means that the probability of surviving will be estimated from a logistic regression adjusted for A, B and C and treatment group. From each treatment group we can do an LSMEANS and get an 'overall' β_1 along with the standard error of β_1.

For a given treatment group the mean probability of surviving is given as:

$$\frac{e^{\beta_1}}{1 + e^{\beta_1}}$$

The distribution of days on vasopressors and mechanical ventilation (30- PVFD's) for those surviving will be estimated from a negative binomial regression (poisson with a potential overdispersion), also adjusted for A, B and C and treatment group. From each treatment group we can do an LSMEANS and get an 'overall' β_2 along with the standard error of β_2.

For a given treatment group the mean PVFD's for survivors is given as:

$$(30 - e^{\beta_2})$$

The mean PVFD's for all subjects (survivors and non-survivors) is then given as:

$$\frac{e^{\beta_1}}{1 + e^{\beta_1}} * (30 - e^{\beta_2})$$

Let

$$f(\beta_1, \beta_2) = \frac{e^{\beta_1}}{1 + e^{\beta_1}} * (30 - e^{\beta_2})$$

Using the delta method we can calculate the variance of f(β_1, β_2):

$$f'(\beta_1,\beta_2) = \left(\left(30 - e^{\beta_2}\right) * \frac{e^{\beta_1}}{(1+e^{\beta_1})^2} , \frac{e^{\beta_1}}{(1+e^{\beta_1})} * (-e^{\beta_2}) \right)$$

Hence,

$$var[f(\beta_1,\beta_2)] = f'(\beta_1,\beta_2) \begin{pmatrix} \sigma^2_{\beta_1} & 0 \\ 0 & \sigma^2_{\beta_2} \end{pmatrix} f'(\beta_1,\beta_2)^t$$

$$= \left(\sigma^2_{\beta_1} * \left(30 - e^{\beta_2}\right) * \frac{e^{\beta_1}}{(1+e^{\beta_1})^2} \quad \sigma^2_{\beta_2} * \frac{e^{\beta_1}}{(1+e^{\beta_1})} * (-e^{\beta_2}) \right) \begin{pmatrix} \left(30 - e^{\beta_2}\right) * \frac{e^{\beta_1}}{(1+e^{\beta_1})^2} \\ \frac{e^{\beta_1}}{(1+e^{\beta_1})} * (-e^{\beta_2}) \end{pmatrix}$$

$$= \sigma^2_{\beta_1} * \left(30 - e^{\beta_2}\right)^2 * \left(\frac{e^{\beta_1}}{(1+e^{\beta_1})^2} \right)^2 + \sigma^2_{\beta_2} * \left(\frac{e^{\beta_1}}{(1+e^{\beta_1})} \right)^2 * e^{2\beta_2}$$

The 95% CI for the mean PVFD's then become

$$\frac{e^{\widehat{\beta_1}}}{1+e^{\widehat{\beta_1}}} * \left(30 - e^{\widehat{\beta_2}}\right) \pm 1.96 * \sqrt{var[f(\widehat{\beta_1},\widehat{\beta_2})]}$$

Let s indicate selepressin and p indicate placebo, the estimated treatment difference then becomes,

$$d = \frac{e^{\widehat{\beta_{1s}}}}{1+e^{\widehat{\beta_{1s}}}} * \left(30 - e^{\widehat{\beta_{2s}}}\right) - \frac{e^{\widehat{\beta_{1p}}}}{1+e^{\widehat{\beta_{1p}}}} * \left(30 - e^{\widehat{\beta_{2p}}}\right)$$

With 95% CI,

$$\frac{e^{\widehat{\beta_{1s}}}}{1+e^{\widehat{\beta_{1s}}}} * \left(30 - e^{\widehat{\beta_{2s}}}\right) - \frac{e^{\widehat{\beta_{1p}}}}{1+e^{\widehat{\beta_{1p}}}} * \left(30 - e^{\widehat{\beta_{2p}}}\right) \pm 1.96 * \sqrt{var[d]}$$

Where $var[d]$ is derived again using the delta method as,

$$var\left[d(\beta_{1s},\beta_{1p},\beta_{2s},\beta_{2p})\right] = d' \begin{pmatrix} \sigma^2_{\beta_{1s}} & \sigma^2_{\beta_{1s}\beta_{1p}} & 0 & 0 \\ \sigma^2_{\beta_{1s}\beta_{1p}} & \sigma^2_{\beta_{1p}} & 0 & 0 \\ 0 & 0 & \sigma^2_{\beta_{2s}} & \sigma^2_{\beta_{2s}\beta_{2p}} \\ 0 & 0 & \sigma^2_{\beta_{2s}\beta_{2p}} & \sigma^2_{\beta_{2s}} \end{pmatrix} d'^t$$

With,

$$d'(\beta_{1s},\beta_{1p},\beta_{2s},\beta_{2p})^t = \begin{pmatrix} \left(30 - e^{\beta_{2s}}\right) * \dfrac{e^{\beta_{1s}}}{(1+e^{\beta_{1s}})^2} \\ -\left(30 - e^{\beta_{2p}}\right) * \dfrac{e^{\beta_{1p}}}{(1+e^{\beta_{1p}})^2} \\ -\dfrac{e^{\beta_{1s}}}{(1+e^{\beta_{1s}})} * e^{\beta_{2s}} \\ \dfrac{e^{\beta_{1p}}}{\left(1+e^{\beta_{1p}}\right)} * e^{\beta_{2p}} \end{pmatrix}$$

In the above reasoning, it is assumed that P&VFDs can only take on integer values between 0 and 30. In practice, P&VFDs will be analysed with one decimal (with values in the range of 0, 0.1, ..., 29.9, 30.0) and hence the distribution of days on vasopressors and mechanical ventilation for those surviving will be estimated from a negative binomial regression model scaled up to 300 (transforming the P&VFDs from 0.0, 0.1, ..., 29.9, 30.0 to 0, 1, ..., 299, 300), and later scaled back to one decimal.

Let X = 10*Y (scaling Y from 0 to 300) and δ_1 and δ_2 the corresponding estimated parameters from the logistic regression and negative binomial regression models based on the modeling of X.

Transforming back to Y, the 95% CI for the mean P&VFDs (E(Y)) then become

$$\frac{e^{\widehat{\delta_1}}}{1+e^{\widehat{\delta_1}}} * \left(30 - \frac{e^{\widehat{\delta_2}}}{10}\right) \pm 1.96 * \sqrt{\frac{var[f(\widehat{\delta_1},\widehat{\delta_2})]}{100}}$$

Again with s indicating selepressin and p indicating placebo, the estimated treatment difference then becomes,

$$\frac{e^{\widehat{\delta_{1s}}}}{1+e^{\widehat{\delta_{1s}}}} * \left(30 - \frac{e^{\widehat{\delta_{2s}}}}{10}\right) - \frac{e^{\widehat{\delta_{1p}}}}{1+e^{\widehat{\delta_{1p}}}} * \left(30 - \frac{e^{\widehat{\delta_{2p}}}}{10}\right)$$

With 95% CI,

$$\frac{e^{\widehat{\delta_{1s}}}}{1+e^{\widehat{\delta_{1s}}}} * \left(30 - \frac{e^{\widehat{\delta_{2s}}}}{10}\right) - \frac{e^{\widehat{\delta_{1p}}}}{1+e^{\widehat{\delta_{1p}}}} * \left(30 - \frac{e^{\widehat{\delta_{2p}}}}{10}\right) \pm 1.96 * \sqrt{\frac{var\left[f\left(\widehat{\delta_{1s}},\widehat{\delta_{2s}},\widehat{\delta_{1p}},\widehat{\delta_{2p}}\right)\right]}{100}}$$

Appendix 4 Estimation of Treatment Effects for Difference of Proportions in Incidence of RRT, and Difference in Number of New Organ Failures and Number of New Organ Dysfunctions

Estimation of Treatment Effects for Difference of Proportions in Incidence of RRT

Let Y be the incidence of RRT (and/or death).

Y can then be modelled using a logistic regression, adjusted for various factors and covariates.

In practice, and with the model adjusted for A, B and C, this means that the probability of RRT incidence will be estimated from a logistic regression (PROC GENMOD preferred over PROC LOGISTIC as estimates can be subtracted by ODS output) adjusted for A, B and C and treatment group. From each treatment group we can do an LSMEANS and get an 'overall'(adjusted) β along with the standard error of β and the covariance between the estimates from each treatment group. I.e. we get β_s, β_p, σ_{β_s},σ_{β_p}, and $\sigma^2_{\beta_s\beta_p}$.

For a given treatment group the mean probability of RRT incidence is given as:

$$\frac{e^{\beta}}{1+e^{\beta}}$$

The difference in proportions of RRT incidence is then given as:

$$\frac{e^{\beta_s}}{1+e^{\beta_s}} - \frac{e^{\beta_p}}{1+e^{\beta_p}}$$

Using the delta method we can calculate the variance of f(β_s, β_p):

$$f'(\beta_s, \beta_p) = \left(\frac{e^{\beta_s}}{(1+e^{\beta_s})^2} , -\frac{e^{\beta_p}}{(1+e^{\beta_p})^2} \right)$$

Hence,

$$var[f(\beta_s, \beta_p)] = f'(\beta_s, \beta_p) \begin{pmatrix} \sigma^2_{\beta_s} & \sigma^2_{\beta_s\beta_p} \\ \sigma^2_{\beta_s\beta_p} & \sigma^2_{\beta_p} \end{pmatrix} f'(\beta_s, \beta_p)^t$$

$$= \left(\frac{\sigma^2_{\beta_s} * e^{\beta_s}}{(1+e^{\beta_s})^2} - \frac{\sigma^2_{\beta_s\beta_p} * e^{\beta_p}}{(1+e^{\beta_p})^2}, \quad \frac{-\sigma^2_{\beta_p} * e^{\beta_p}}{(1+e^{\beta_p})^2} + \frac{\sigma^2_{\beta_s\beta_p} * e^{\beta_s}}{(1+e^{\beta_s})^2} \right) \begin{pmatrix} \frac{e^{\beta_s}}{(1+e^{\beta_s})^2} \\ \frac{-e^{\beta_p}}{(1+e^{\beta_p})^2} \end{pmatrix}$$

$$= \sigma^2_{\beta_s} * \left(\frac{e^{\beta_s}}{(1+e^{\beta_s})^2} \right)^2 + \sigma^2_{\beta_p} * \left(\frac{e^{\beta_p}}{(1+e^{\beta_p})^2} \right)^2 - 2 * \sigma^2_{\beta_s\beta_p} * \frac{e^{\beta_s}}{(1+e^{\beta_s})^2} * \frac{e^{\beta_p}}{(1+e^{\beta_p})^2}$$

The 95% CI for the difference in proportions of RRT incidence then become

$$\frac{e^{\widehat{\beta_s}}}{1+e^{\widehat{\beta_s}}} - \frac{e^{\widehat{\beta_p}}}{1+e^{\widehat{\beta_p}}} \pm 1.96 * \sqrt{var\left[f\left(\widehat{\beta_s}, \widehat{\beta_p}\right)\right]}$$

Estimation of Treatment Effects for Difference in Number of New Organ Failures and Number of New Organ Dysfunctions

Let Y be the number of new organ dysfunctions (or organ failures).

Y can then be modelled using a negative binomial model (poisson model with possible overdispersion), adjusted for various factors and covariates.

In practice, and with the model adjusted for A, B and C, this means that the number of new organ failures will be estimated from a negative binomial model (PROC GENMOD) adjusted for A, B and C and treatment group. From each treatment group we can do an LSMEANS and get an 'overall'(adjusted) β along with the standard error of β and the covariance between the estimates from each treatment group. I.e. we get β_s, β_p, σ_{β_s},σ_{β_p}, and $\sigma^2_{\beta_s\beta_p}$.

For a given treatment group the mean number of new organ dysfunctions is given as:

$$e^{\beta}$$

The difference in the number of new organ dysfunctions is then given as:

$$e^{\beta_s} - e^{\beta_p}$$

Using the delta method we can calculate the variance of f(β_s, β_p):

$$f'\left(\beta_s, \beta_p\right) = \left(e^{\beta_s}, -e^{\beta_p}\right)$$

Hence,

$$var[f\left(\beta_s, \beta_p\right)] = f'\left(\beta_s, \beta_p\right)\begin{pmatrix} \sigma^2_{\beta_s} & \sigma^2_{\beta_s\beta_p} \\ \sigma^2_{\beta_s\beta_p} & \sigma^2_{\beta_p} \end{pmatrix} f'\left(\beta_s, \beta_p\right)^t$$

$$= \left(\sigma^2_{\beta_s} * e^{\beta_s} - \sigma^2_{\beta_s\beta_p} * e^{\beta_p},\quad \sigma^2_{\beta_s\beta_p} * e^{\beta_s} - \sigma^2_{\beta_p} * e^{\beta_p}\right)\begin{pmatrix} e^{\beta_s} \\ -e^{\beta_p} \end{pmatrix}$$

$$= \sigma^2_{\beta_s} * e^{2\beta_s} + \sigma^2_{\beta_p} * e^{2\beta_p} - 2 * \sigma^2_{\beta_s\beta_p} * e^{\beta_s} * e^{\beta_p}$$

The 95% CI for the difference in the number of new organ dysfunctions then become

$$e^{\widehat{\beta_s}} - e^{\widehat{\beta_p}} \pm 1.96 * \sqrt{var\left[f\left(\widehat{\beta_s}, \widehat{\beta_p}\right)\right]}$$

Appendix 5 Statistical Model for Adaptive Design Decisions

An Example Trial

This section presents the results of an example trial in order to illustrate how the design behaves. The selected scenario includes defining the probability distributions that describe how patients given placebo behave, and also defining the effects of all four selepressin dosing regimens. In the example trial we work with P&VFD data recorded to the nearest day instead of the nearest tenth of a day.

Figure 6 shows the data available at the first interim analysis, and the results of the statistical analyses performed using those data that are then used to make decisions. The leftmost of the three plots shows the raw data for each of the five arms, with mortality data in the legend and P&VFD data for survivors in the barplots. Red represents the placebo arm, and the shades of blue and black represent the active arms, with darker colors indicating larger maximum doses. The curves added to the plot are naïve density estimates that are not related to the statistical models used in the trial, scaled so that they have the same maximum. Active arm 2 has observed the highest mortality rate, while active arms 1 and 3 have seen tentative improvements compared to placebo. Active arm 3 has had the more surviving patients with small numbers of P&VFDs than the other arms. Active arm 4 is not yet allowed to accept patients. The middle plot displays estimates of mean P&VFDs (with fatalities scored as zero P&VFDs), together with 95% uncertainty intervals. The placebo arm is estimated to be the least effective, and uncertainties are smallest for the intermediate doses. The rightmost plot displays predictive probabilities and updated allocation probabilities. The circles show the allocation probabilities for the next 30 days: the probability for placebo remains at 1/3 for the duration of Part 1, and active arm 3 is assigned the lowest probability for any active arm. The rightmost plot also features squares showing the predictive probability that this trial would be successful if Part 1 were to terminate and Part 2 were to begin with each of the four arms as the chosen arm (It is too early to choose an arm: this can only happen when final data have been observed for at least 300 patients). If all four active arms had predictive probabilities less than 0.05 the trial would terminate for futility. Also, if the predictive probability were at least 90% that the final data set would show an observed increase in mortality of greater than 2%, regardless of which of the four active arms were selected, the trial would stop for futility for that reason, but these predictive probabilities are not shown in the figures.

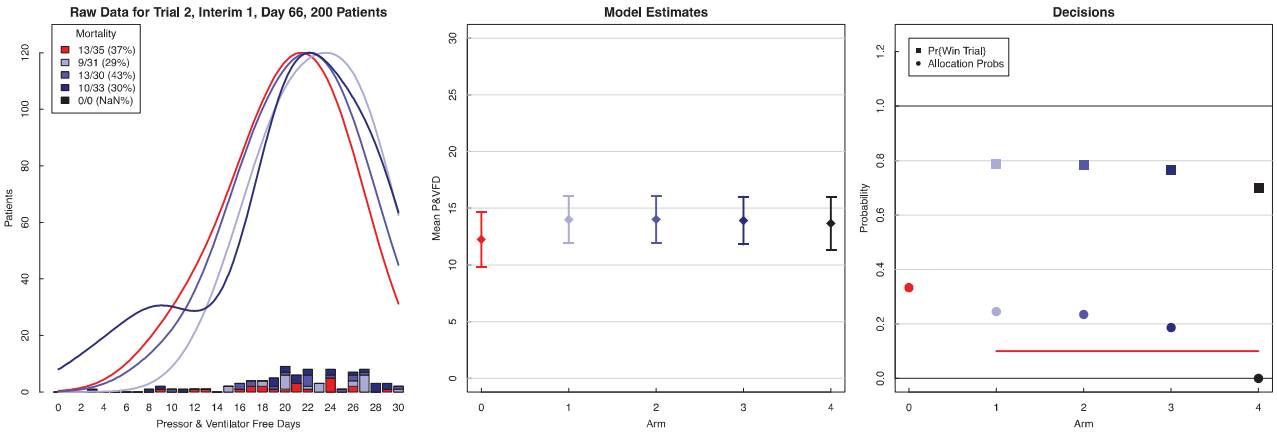

Figure 6: Data and interim analysis results at time of first interim analysis in an example trial

Moving ahead to the second interim analysis shown in Figure 7, when we have observed final data for 199 patients. Mortality has remained high for active arm 2. For survivors, the frequency of small numbers of P&VFDs is lowest for the placebo arm, but all three active arms have seen more patients with very large numbers of P&VFDs than has placebo. According to the rightmost plot, predictive probabilities for trial success are around 40% for all active arms. For the next 30 days, arm 3 will receive the highest allocation probability among the active arms, and since (not shown) the probability is at least 50% that arm 3's expected P&VFDs is higher than for arm 2, arm 4 becomes eligible for patient allocation to explore whether the apparent increasing trend in efficacy continues.

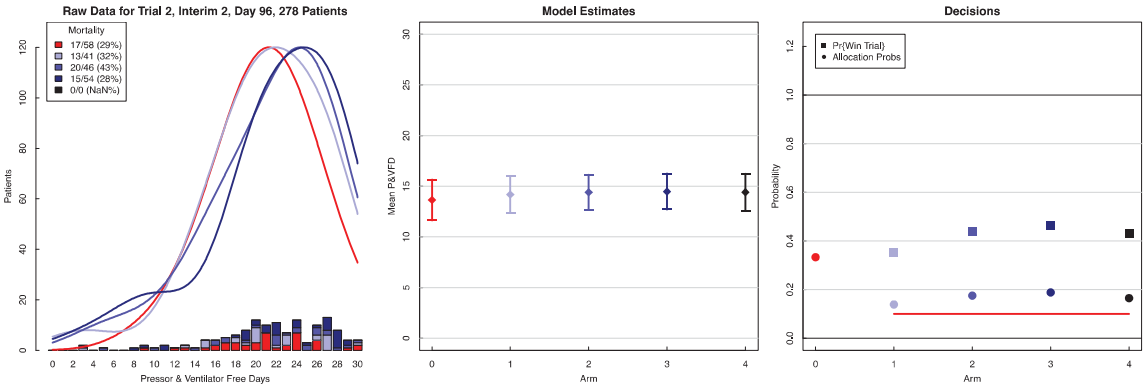

Figure 7: Data and analysis results for second interim analysis in example trial.

We skip ahead to the fourth interim analysis, shown in Figure 8, which is the first interim analysis after final data are available for at least 300 patients, so this is the first opportunity to choose a dose and move on to Part 2. The first data from arm 4 have come in, and arms 1, 3, and 4 are all achieving mortality rates no worse than placebo. The placebo arm continues to have the smallest probability of large numbers of P&VFDs among survivors. Since the predictive probability of trial

success is largest for arm 3 and its value exceeds 90% (shown by the light green line), the trial elects to terminate Part 1 and proceed to Part 2 with arm 3. The allocation probabilities change to 50% each for placebo and active arm 3. Unless the trial stops earlier for futility, Part 2 will consist of 1430 patients since 370 are in the trial currently. The decision to select arm 3 occurs with only 11 patients allocated to arm 4, and with arm 4 looking promising thus far, but arm 3 has a high enough probability of a successful trial that the trial advances to Part 2.

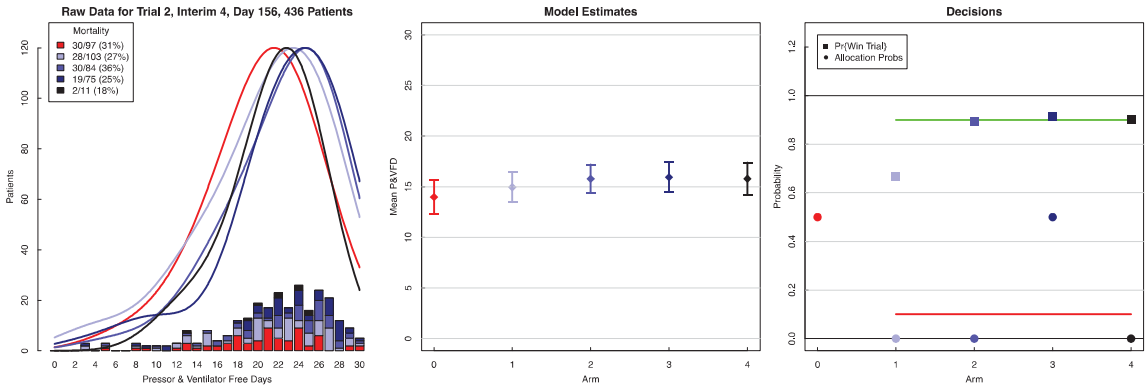

Figure 8: Data and analysis results for fourth interim analysis in example trial

During this interim analysis, Part 1 terminates and the trial proceeds to Part 2 with active arm #3.

The first interim analysis of Part 2 takes place after 63 more patients are allocated to arm 3 or placebo. Final data continue to come in for the other active arms as well. Since data are relatively favourable for the placebo arm in this month, the predictive probability of a successful trial at 1800 patients drops to about 0.80, which is much larger than the 5% futility boundary for Part 2 of the trial.

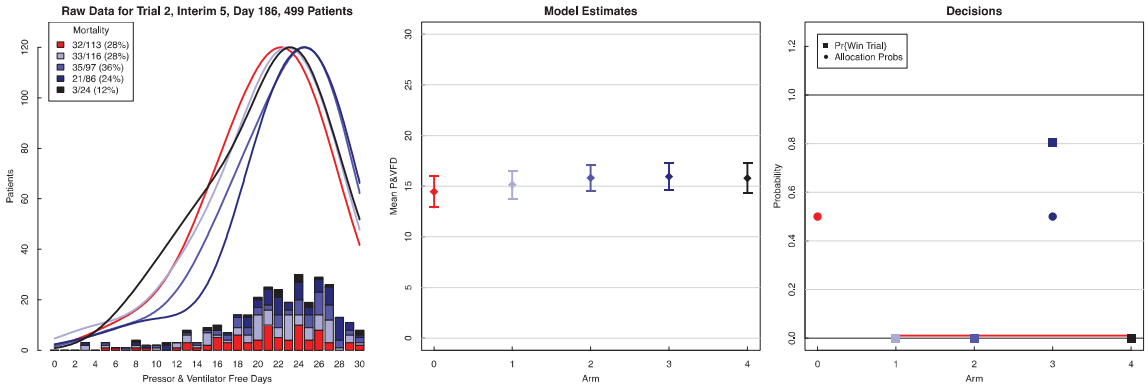

Figure 9: First interim analysis during Part 2

Interim analyses continue every 30 days (frequency chosen for this example), with the only available decision being whether or not to stop for futility. The predictive probability of trial success never approaches the 5% value that would trigger a futility stop. The final data at the end of

the trial are shown in Figure 9. Observed mortality was 2% lower for active arm 3 than for placebo, and the final analysis will be based on a smaller difference than this due to the inclusion of the data from other active arms. The simulation assumed true mortalities and expected P&VFD distributions for survivors are shown in the middle plot: the doses increase in effectiveness at both preventing mortality and increasing P&VFDs for survivors, with arms 3 and 4 each reducing mortality by 1.5% and increasing expected P&VFDs by 1.5 days for survivors. The rightmost plot shows the final p-value for the Wilcoxon test: the p-value for the comparison between patients given placebo and patients given any active arm is lower than 0.025, so this is a successful trial.

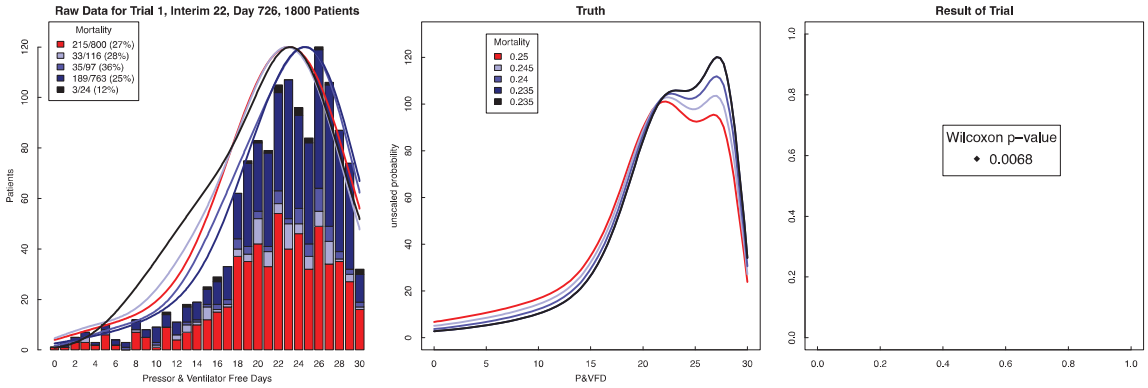

Figure 10: Results of final analysis. The trial is successful, with a p-value of 0.0068.

Statistical Model for Adaptive Design Decisions

A patient's outcome can be either death or a number of vasopressor and mechanical ventilator-free tenths of days between 0 and 300. Label d_i as the dose regimen assigned to patient i, and denote by s_i the stratum to which the patient belongs ($s_i \in \{0,1,\ldots,7\}$). First, define the distribution of these potential outcomes for patients treated with placebo ($d_i = 0$). If the i'th patient dies, write $X_i = 1$, otherwise $X_i = 0$. So,

$$\Pr\{X_i = 1 | \, d_i = 0, s_i = 0\} = \Delta.$$

Intuitive parameterization of the P&VFD model

To motivate the strategy for modeling the distribution of the number of P&VFD given placebo and survival, we first define a version of the model that is more intuitive but less computationally convenient and is asymmetric with respect to the strata. The version actually recommended is defined below in the section "*Symmetric, computational parameterization of the P&VFD model* ". The distribution can be modeled nonparametrically:

$$\Pr\{Y_i = k | d_i = 0, s_i = 0, X_i = 0\} = \pi_k, \text{ with } \sum_{j=0}^{300} \pi_j = 1.$$

Note that patients who die are modelled separately from patients who survive, but nevertheless accumulate zero P&VFD, although these patients are handled together in the final analysis. We model the treatment effect for a given dose d with two parts: the effect on mortality ϕ_d and the

effect on P&VFD given survival, θ_d. The differences between the strata are modeled similarly to the effects of the different doses, with a stratum effect on mortality denoted by ψ_s and a stratum effect on P&VFD given survival denoted by ω_s. By definition $\psi_0 = \omega_0 = 0$. We model the effect on mortality, ϕ_d, on the log-odds scale:

$$\Pr\{X_i = 1 | d_i = d, s_i = s\} = \frac{\Delta}{\{\Delta + (1-\Delta)\exp(\phi_d + \psi_s)\}},$$

where we have defined the effect so that larger values of ϕ_d or ψ_s are beneficial (decrease mortality). We model the effects of the dose arm and the stratum on the number of P&VFDs for survivors using an exponential family whose sufficient statistic is the number of P&VFD:

$$\Pr\{Y_i = k | d_i = d, s_i = s, X_i = 0\} = \frac{\pi_k e^{k(\theta_d + \omega_s)}}{\sum_{j=0}^{300} \pi_j e^{j(\theta_d + \omega_s)}}.$$

In particular, given a dose arm, a sample of P&VFD, and a probability vector π, the maximum likelihood estimators of θ_d and ω_s set the expected values of P&VFD equal to their sample means.

In the final analysis, however, patients who died are treated in the same way as patients who survived but who had 0 P&VFD.

For the simpler statistical model with no consideration of stratum effects, the ψ_s and ω_s are omitted.

Symmetric, computational parameterization of the P&VFD model

The version of the statistical model we will actually use is as follows. There are still parameters $\pi_0, \ldots, \pi_{300}$, but they do not correspond directly to any stratum and do not have a clear interpretation of their own. Instead of assuming that $\psi_0 = \omega_0 = 0$, we use another choice of identifiability assumptions: $\sum_{s=1}^{8} \psi_s = 0$ and $\sum_{s=1}^{8} \omega_s = 0$. Now we have

$$\Pr\{Y_i = k | d_i = d, s_i = s, X_i = 0\} = \frac{\pi_k e^{k(\theta_d + \omega_s)}}{\sum_{j=0}^{300} \pi_j e^{j(\theta_d + \omega_s)}}$$

for all d and s, where we have defined the placebo parameter $\theta_0 = 0$. This parameterization facilitates the definition of prior distributions for the ψ_s and ω_s, and ensures that the results of the analysis do not depend on which stratum is identified with $s = 0$, etc. We assume hierarchical models on the stratum parameters, with $\psi_s \sim N(0, \tau_\psi^2)$ and $\omega_s \sim N(0, \tau_\omega^2)$ (conditionally on observing the identifiability assumptions), and with $\tau_\psi \sim$Uniform(0,1) and $\tau_\omega \sim$Uniform(0,1).

Inverted-U dose-response models for the dose effects ϕ and θ.

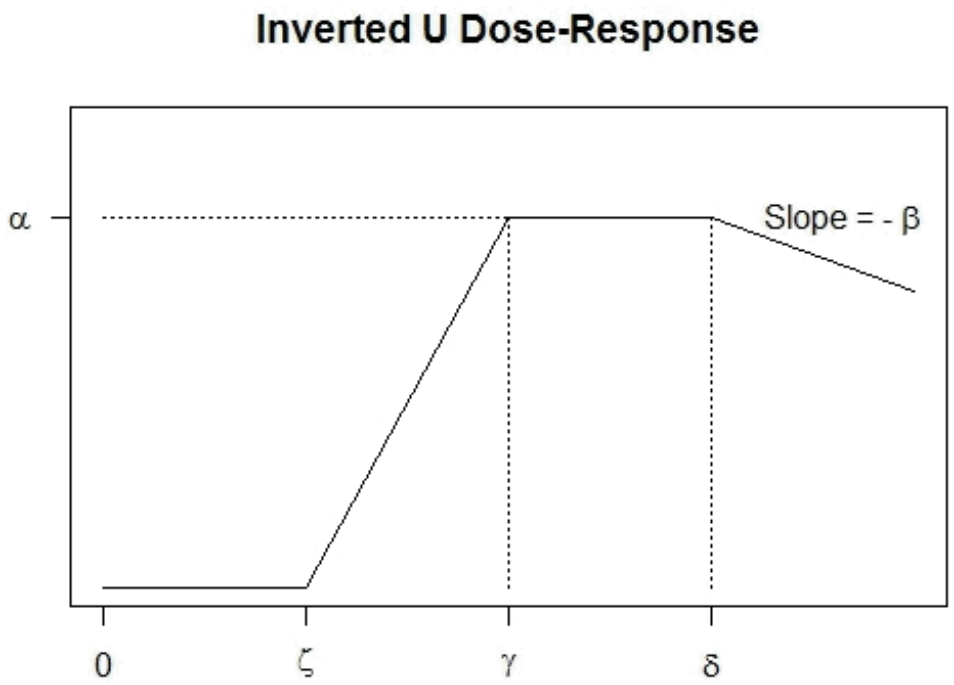

Figure 11: Inverted-U dose-response model

The x-axis represents dose regimen (which is interpreted as maximum dose in the selepressin trial), and the y-axis represents the treatment effect. The model parameters ζ, γ, δ define theoretical special doses (e.g. γ is the smallest dose that gives maximal effect), while α denotes the maximum achievable treatment effect.

An inverted-U dose-response model is used for the effects on both mortality and P&VFD. This is a flexible family that allows for the possibilities that small doses have no effect at all (if those doses are smaller than ζ), and that large doses can start to become less effective than smaller doses (if those doses are larger than δ). Most importantly for this trial, the model allows all doses to be equally effective, since, for a given patient, any assigned maximum dose could be titrated to a similar delivered treatment. In the parameterization, shown in Figure 10, γ is the smallest dose that delivers maximal effect, δ is the largest dose that delivers maximal effect, ζ is the largest dose that delivers no effect, β is the rate at which performance degrades beyond dose δ, and α is the size of the largest effect ($\alpha \leq 0$ is also allowed and there is a prior distribution for α which is symmetric about zero. To maintain an inverted U shape, β is constrained to be positive). Care must be taken to avoid important identifiability problems: in particular, it is insisted that the interval (γ, δ) contains at least one dose for which data can potentially be obtained, otherwise the largest treatment effect is not uniquely defined. However, γ can be smaller than any active dose, in which case the data contain no information about ζ, and δ can be larger than any active dose, in which case the data contain no information about β, but neither of these nonidentifiabilities are serious problems since they do not impact the ability to predict future data.

In the case where the candidate doses are 1, 2, 3, and 4, a prior distribution is assumed in which

$$\delta - \gamma \sim \text{Uniform}(1,5),$$

$$\gamma|\delta \sim \text{Uniform}(0, \delta - 1), \text{ and}$$

$$\zeta|\gamma \sim \text{Uniform}(0, \gamma).$$

It is assumed that $\phi_d = IU(d|\alpha_1, \zeta_1, \gamma_1, \delta_1, \beta_1)$ and $\theta_d = IU(d|\alpha_2, \zeta_2, \gamma_2, \delta_2, \beta_2)$ for $d = 1,2,3,4$, where the inverted-U function is defined as

$$IU(d|\alpha, \zeta, \gamma, \delta, \beta) = \alpha\{(1 - \left[1 - \frac{(d-\zeta)_+}{(\gamma-\zeta)}\right]_+\Big)_+ - \beta(d-\delta)_+\}, \text{ where } x_+ = \max(x, 0).$$

There is no assumed relationship between the θ parameters and the ϕ parameters.

For a specific example, suppose $\alpha_1 = 0.2, \zeta_1 = 1.5, \gamma_1 = 2.5$, and $\delta_1 = 4$ (so that the value of β_1 is irrelevant for doses 1, 2, 3, and 4). Then $\phi_1 = 0, \phi_2 = 0.1, \phi_3 = \phi_4 = 0.2$.

To complete the specification of the prior distribution, it is assumed that

$$\Delta \sim Beta\left(\frac{1}{3}, \frac{1}{3}\right), (\pi_0, \pi_1, \dots, \pi_{300}) \sim \text{Dirichlet}\left(\frac{1}{30}, \frac{1}{30}, \dots, \frac{1}{30}\right),$$

$$\alpha_1 \sim \text{Uniform}(-0.35, 0.35),$$

$$\alpha_2 \sim \text{Uniform}(-0.01334, 0.01334),$$

$$\beta_1 \sim \text{Uniform}(0,10), \text{ and}$$

$$\beta_2 \sim \text{Uniform}(0,1).$$

The limits for the α parameters are selected to be large but not completely absurd; e.g. $\alpha_1 = 0.35$ corresponds to a 5% benefit to a 20% mortality rate. Given P&VFD data, one can use a variable-at-a-time Metropolis-Hastings algorithm to sample from the posterior distribution of the unknown parameters $(\Delta, \pi_0, \pi_1, \dots, \pi_{300}, \alpha_1, \zeta_1, \gamma_1, \delta_1, \beta_1, \alpha_2, \zeta_2, \gamma_2, \delta_2, \beta_2)$. These samples can be used to estimate the predictive probability of a successful Part 2, the probability that each doses maximizes the expected number of P&VFDs, and the probability that dose regimen 3 provides a larger value of expected P&VFDs than does dose regimen 2.

The scale of α_2 depends on how P&VFDs are measured. Here they are recorded as integer numbers of tenths. If integer days are used instead, the prior range should be expanded by a factor of ten. The Dirichlet prior exponents we use also depend on how P&VFDs are measured: with integer days, we use exponents of 1/3 instead of 1/30. In either case the prior distribution corresponds to approximately ten equivalent observations.

Response-Adaptive Randomisation Probabilities

Beginning with the first interim analysis, allocation probabilities for the selepressin arms are adjusted based on the posterior distribution (the allocation probability for placebo remains fixed at 1/3 throughout Part 1). The allocation probability for active arm *j* is proportional to the posterior probability that arm *j* is one of the arms that maximizes the expected number of P&VFD (where the mortality probabilities are included in the calculations). Note that the inverted-U models allow for multiple arms to have exactly the same expected number of P&VFD.

Posterior Predictive Probability Calculations

Computations of the predictive probability of a successful Part 2 are critical to the design. These computations proceed by drawing samples from the posterior distribution of the unknown parameters. We first discuss the predictive probability calculation for the simplified case in which there are no strata and the final analysis is a Wilcoxon test; this is the case that applies in the operating characteristics simulations presented in this report. For a given posterior sample, we use a normal approximation to the predictive distribution of the Wilcoxon statistic that will be obtained from the currently available data for placebo and the pooled active arms. The calculation is tedious but straightforward, based on writing the Wilcoxon statistic in terms of

$$\sum_{i=1}^{I}\sum_{j=1}^{J} sign\left(Y_{1i} - Y_{2j}\right),$$

where $Y_{11}, \ldots, Y_{1I}$ are the P&VFDs for sample 1 (i.e. the active arm) and $Y_{21}, \ldots, Y_{2J}$ are the P&VFDs for sample 2 (i.e. the placebo arm). For a given set of P&VFD probabilities for the two arms, one can calculate the expected value and variance of the Wilcoxon statistic, and use this to calculate the probability of a significant result (the variance depends on the number of ties, and we plug in the expected value of the tie component based on a Poisson approximation). The process is then repeated for more posterior samples, and the results are averaged to give the overall estimate of the predictive probability of a successful trial. The predictive probability calculation is performed for each active arm being the one that proceeds to Part 2. The calculation uses the same (pooled) data set for each arm, but each arm has a different posterior distribution of treatment effect and hence a different predictive distribution of future data.

The approximate formula for predictive probability of a successful Wilcoxon test is faster computationally than simulating many final data sets and calculating whether each one attains success, so it plays a key role in simulating the trial to estimate operating characteristics. When time permits, such as when the design is being executed, however, the direct Monte Carlo simulation should also be performed.

The extension to the case in which there are eight strata and in which the final analysis is a van Elteren test, is straightforward but adds another layer of complexity. For the purpose of this calculation, we introduce a Dirichlet-multinomial model for the stratum probabilities: the eight stratum probabilities begin with a Dirichlet prior distribution with parameters equal to 1/3 and then

are updated with the stratum counts observed in the trial. This stratum probability model operates independently of the remainder of the statistical modeling. Suppose that we are interested in calculating the predictive probability of success assuming that half the remaining patients are allocated to the placebo and the other half are allocated to active arm 1. For a given posterior sample, we draw a sample from the Dirichlet posterior distribution of the stratum, and then draw multinomial samples for the stratum counts for the future patients allocated to placebo, and separately to the active arm. We then loop over the strata and either

- using the posterior sample, simulate numbers of deaths and P&VFDs for survivors for each arm and each stratum and then evaluate the Wilcoxon test statistic, its theoretical null hypothesis mean, and its theoretical null hypothesis variance (which depends on the numbers of ties in the data set), or
- based on the posterior sample and the current numbers of deaths and P&VFD counts in each arm and the current stratum, calculate the predictive mean and variance of the Wilcoxon statistic, its theoretical null hypothesis mean, and an approximation to the expected value of the theoretical null hypothesis variance. The current data for all the active arms are aggregated together.

Denoting the Wilcoxon test statistic for the sth stratum by T_s, the simulated final number of patients in stratum s by N_s, and the null hypothesis expected value and variance by $E(T_s|H_0)$ and $Var(T_s|H_0)$ respectively, the van Elteren test statistic is

$$T = \frac{\sum_{s=1}^{8}\{T_s - E(T_s|H_0)\}/(N_s+1)}{\{\sum_{s=1}^{8} Var(T_s|H_0)/(N_s+1)^2\}^{1/2}}$$

For the Monte Carlo estimate of predictive probability, evaluate this statistic for every simulated final data set and compute the fraction of final data sets for which this statistic exceeded the 97.5th Gaussian percentile. To use the approximate formula, denote the predictive mean of T_s by $E(T_s|D,i)$, the predictive variance by $Var(T_s|D,i)$, and the estimated null hypothesis variance by $E\{Var(T_s|H_0)|D,i)\}$; D denotes the current data and i denotes that we are using the ith posterior sample. The approximate predictive probability for posterior sample i is then computed using Gaussian tail probabilities based on the expected value and variance of T given by

$$E(T|i) = \frac{\sum_{s=1}^{8}\{E(T_s|D,i) - E(T_s|H_0)\}/(N_s+1)}{\{\sum_{s=1}^{8} E\{Var(T_s|H_0)|D,i)\}/(N_s+1)^2\}^{1/2}}$$

$$Var(T|i) = \frac{\sum_{s=1}^{8}\{Var(T_s|D,i)\}/(N_s+1)^2}{\sum_{s=1}^{8} E\{Var(T_s|H_0)|D,i)\}/(N_s+1)^2}$$

Similar but simpler calculations apply for estimating the predictive probability that the final data set will have a mortality rate among the patients assigned to an active arm that is at least 2% higher than among the placebo patients.

Control of Type I Error

This design achieves control of Type I error through analytical means. While the trial can stop early for futility, a successful trial can only be achieved at a total of 1800 patients. At this time, a single test statistic is calculated, and it compares two populations that are defined before the trial begins, namely patients allocated to placebo compared to patients that are allocated to any active arm. In particular, no patients are excluded from the final analysis for any reason related to their outcomes (in contrast, if the final analysis compared placebo to the best performing of the active arms, that would inflate Type I error).

This argument demonstrates that Type I error is controlled even for the modification of the design in which early stopping for futility is disabled (See Appendix 6 for a formal proof of Type I error control). The potential for early stopping for futility, including the 25% predictive probability requirement at 800 patients, further reduces Type I error probability below the nominal value.

Appendix 6 Formal Proof of Type I Error Control

The following assumptions are made in the statistical discussions of type I error.

1. The primary analysis is based on a Wilcoxon-Mann-Whitney test (Van Elteren's Test) on the P&VFDs. The normal approximation to the sampling distribution of the test statistic is assumed for type I error discussions (as it would in a fixed trial).
2. The primary analysis combines all the active arms together for the final analysis. Under the null hypothesis, all arms (placebo and active) have the same mean and therefore the active arms can be combined in to a single arm.
3. The final analysis is based on 1800 patients.
4. The only "adaptive" aspect of the trial is the time in which randomisation switches from 2:1 (active to placebo) to (1:1).
    a. It is assumed that if the randomisation was 2:1 for the entire 1800 patients that type I error is controlled.
    b. It is assumed that if at a fixed point in time (deterministic) randomisation went from 2:1 to 1:1 that type I error is controlled.

We demonstrate the control of type I error by first considering a one-sample problem. We use the one-sample result to then extend to the two-sample problem. The proof focuses on the notion that when the data are positive in the first part of the trial it triggers a shift to a randomisation that increases the *effective* sample size of the trial. As demonstrated in Mehta and Pocock (2011) (**3**) this controls type I error when the data that triggers the shift are appropriately positive.

The hypothesis test is

$$H_o: \mu = 0$$

$$H_1: \mu > 0$$

Assume iid normally distributed random variables are observed with mean μ and known variance of 1. At the interim time point a random variable, based on n_0 observations, $Z_0 = n_0\bar{X}_0$ is observed. After this time point two different random variables (with entirely different observations, X) could be observed: $Z_1 = n_1\bar{X}_1$ or $Z_2 = n_2\bar{X}_2$. We use the notation N(μ,σ^2) for a normal distribution with mean μ and variance σ^2.

So, under the null hypotheses,

$$Z_o \sim N(0, n_0)$$
$$Z_1 \sim N(0, n_1)$$

and,

$$Z_2 \sim N(0, n_2).$$

The adaptive design specifies that if the data are unfavorable we will observe Z_1, and if the data are favorable we will observe Z_2. Therefore, for some value *b*,

1. If the data are unfavorable ($Z_0 < b$) we observe a second random variable $Z_1 \sim N(0, n_1)$. The trial is declared a success at the end of the second stage if

$$Z_0 + Z_1 > a\sqrt{n_0 + n_1}$$

where

$$a = \Phi^{-1}(1 - p).$$

If $b=\infty$ then it is deterministic (select Z_1) and the probability of success is p (the type 1 error of a fixed design, under the null).

2. If the data are favorable ($Z_0 > b$) we observe the random variable $Z_2 \sim N(0, n_2)$. We assume that the sample size for Z_2 is $n_2 = rn_1$. The parameter *r* is a flexible parameter for increase ($r>1$) or decrease ($r<1$) in the sample size in the second part of the trial, so $r>1$ implies $n_2 > n_1$. The trial is declared a success at the end of the second stage if

$$Z_0 + Z_2 > a\sqrt{n_0 + rn_1}$$

The critical value has been set so that if $b=-\infty$ (again deterministic to select Z_2) then the probability of success is p (type I error under a fixed design).

The value of *b* then determines the type I error of the adaptive design. The probability of a type I error of the adaptive design, for *f* the pdf of Z_0, is then

$$\int_{-\infty}^{b} f(z)\,\Phi\left(\frac{z-a\sqrt{n_0+n_1}}{\sqrt{n_1}}\right)dz + \int_{b}^{\infty} f(z)\,\Phi\left(\frac{z-a\sqrt{n_0+rn_1}}{\sqrt{rn_1}}\right)dz \qquad (1)$$

If $r=1$ there is no change in the trial and the above (1) is equal to *p*. In this trial $r > 1$. If the expression in (1) is nonincreasing in *r* then this demonstrates that the type I error is not inflated above *p* in this design. A sufficient condition is if $\frac{z-a\sqrt{n_0+rn_1}}{\sqrt{rn_1}}$ is decreasing in *r*. The derivative of this is

$$\frac{d}{dr}\left(\frac{z-a\sqrt{n_0+rn_1}}{\sqrt{rn_1}}\right) = -\frac{n_1}{2(rn_1)^{\frac{3}{2}}}\left(z-\frac{an_0}{\sqrt{n_0+rn_1}}\right)$$

which is negative in the second integral ($z > b$) if

$$b > \frac{an_0}{\sqrt{n_0+rn_1}}. \qquad (2)$$

Therefore, the type I error of the design is guaranteed if expression (2) holds. It is potentially possible to find a smaller *b* that guarantees a reduction in type I error, but that is not required here (see Mehta and Pocock (2011) for further discussion of this reduction). To interpret the constraint on *b*, note that the MLE of μ based on Z_0 is $\widehat{\mu_0} = \frac{Z_0}{n_0}$, and that the MLE of μ based on Z_0 and Z_2 is $\widehat{\mu_{0+2}} = \frac{Z_0+Z_2}{n_0+n_2}$. To elect to observe Z_2 we require

$$\widehat{\mu_0} = \frac{Z_0}{n_0} \geq \frac{b}{n_0} > \frac{a}{\sqrt{n_0+rn_1}},$$

and to show significance at the end of the trial we require

$$\widehat{\mu_{0+2}} = \frac{Z_0+Z_2}{n_0+n_2} > \frac{a\sqrt{n_0+rn_1}}{n_0+rn_1} = \frac{a}{\sqrt{n_0+rn_1}}.$$

In other words, if we are to elect to observe the larger sample size with Z_2, the observed data at the interim must achieve what would need to be observed at the conclusion of the trial in order to be a successful trial (meeting the definition of 50% conditional power). This interpretation is convenient for discussing the Ferring decision rules.

Two-Sample Extension

The condition above is expressed as a single-sample case. We demonstrate the extension to the two-dimensional case. In the single-sample case the test-statistic based on unit-variance normal $X_1, \ldots, X_M$ is

$$\sum_{m=1}^{M} X_m = M\bar{X}_m,$$

which has variance M. For the two-sample case ($X_1, \ldots, X_M$ and $Y_1, \ldots, Y_N$) the analogous statistic is $\frac{MN}{M+N}(\bar{X}_M - \bar{Y}_N)$, which has variance $\frac{MN}{M+N}$ (functions like the sample size). The joint distribution of $\frac{M_kN_k}{M_k+N_k}\left(\bar{X}_{M_k} - \bar{Y}_{N_k}\right)$ for k=1,2,… where the sample sizes M_k and N_k are nondecreasing in k, is the distribution of Brownian motion with drift $E(X_1) - E(Y_1)$ evaluated at the times $\frac{M_kN_k}{M_k+N_k}$: k=1,2,…, which is exactly the same as the one-sample case.

In the Ferring trial since for fixed $M+N$ the effective sample size $\frac{MN}{M+N}$ is maximized when $M=N$, an earlier shift to 1:1 randomisation is synonymous with an *increase* in sample size.

Decision to "Increase Sample Size"

In the Ferring trial the decision to shift to 1:1 and thus increase the sample size is not based on a conditional power or a point estimate, but rather is based on Bayesian predictive probability. The design will shift to 1:1 randomisation early if the predictive probability is greater than 90% (typically a 50% predictive probability is consistent with observed data at the interim larger than what would need to be seen at the final analysis). This can be demonstrated explicitly in the

Gaussian case with prior distribution centered at no treatment effect. For the special case of a flat, improper prior and unit variances, we have that the predictive probability of success

$$\Pr\left\{\bar{X}_{M_0+M_2} - \bar{Y}_{N_0+N_2} > a\sqrt{\frac{1}{M_0+M_2}+\frac{1}{N_0+N_2}}\,|\bar{X}_{M_0},\bar{Y}_{N_0}\right\}$$

is equal to

$$\Phi\left(\frac{\bar{X}_{M_0}-\bar{Y}_{N_0}-a\sqrt{\frac{1}{M_0+M_2}+\frac{1}{N_0+N_2}}}{\sqrt{\frac{1}{M_2}+\frac{1}{N_2}-\frac{1}{M_0+M_2}-\frac{1}{N_0+N_2}}}\right).$$

If the predictive probability is greater than 50% then the numerator is positive and the point estimate at the interim is larger than what is needed for success at the final analysis. If the prior is not improper, but is conjugate normal with mean zero, then the observed effect at the interim must be even larger for a 50% predictive probability of success. In the Ferring trial the prior is more complicated -- given the joint model over multiple active arms, but each of these arms has prior mean the same as the placebo.

Hence the condition of 90% predictive probability to shift to the 1:1 randomisation provides a conservative type I error for the final analysis.

All simulations have reinforced the conservative nature of the final analysis, accounting for the Wilcoxon final analysis.

Post-Simulation P-value

Next to the p-value based on the asymptotic normality of the Van Elteren test, post-simulation bootstrap p-values will be provided as sensitivity analyses.
To this end, the primary outcome for each patient in the trial, regardless of the treatment arm (restricted to the evaluable patients) will be placed in a vector and used to simulate complete trials and bootstrapped test-statistics.

The following procedure will be used:

1. The adaptive trial will be simulated exactly as designed in the protocol
2. The following assumptions will be used for the simulated trial:
    a. The day of enrollment during the actual trial, for each patient that was evaluable will be noted. These days will be used deterministically for the simulated trials.
    b. Additionally the strata in which that actual patient belonged will be recorded and used deterministically in the trial.
3. The randomization rules will follow the adaptive algorithm and design.
4. All, patients, from any arm, will be simulated with replacement from all patients in the actual trial with the same stratification membership.
5. The final test-statistic (Van Elteren's) will be calculated for each simulated trial.

6. This process will be repeated 100,000 times. The probability of a test statistic more extreme than the observed test-statistic is the empirical estimate of the p-value

A second simulation will run identically, but there will be no futility stopping in the simulated trial.

Appendix 7 Randomisation Plan

Part 1: Burn-In:

During burn-in randomisation will be conducted using blocked randomisation. At each site separate lists will be conducted for each of the following 8 strata:

Strata	Mechanical Ventilation	Creatinine	norepinephrine/ noradrenaline
1	Yes	<150 μmol/L	<30 μg/min
2	Yes	<150 μmol/L	≥30 μg/min
3	Yes	≥150 μmol/L	<30 μg/min
4	Yes	≥150 μmol/L	≥30 μg/min
5	No	<150 μmol/L	<30 μg/min
6	No	<150 μmol/L	≥30 μg/min
7	No	≥150 μmol/L	<30 μg/min
8	No	≥150 μmol/L	≥30 μg/min

Each blocked list will be generated using random block sizes of 9 and 18 with ratios of 3:2:2:2:0 for the placebo and active treatment arms 1, 2, 3, and 4, respectively.

Part 1: Adaptive Randomisation:

After burn-in randomisation will be conducted using a mixture of blocking and response adaptive randomisation (RAR) with stratification weighting.

A blocking system will be created for placebo and active arms within each of the 8 strata. Using random block sizes of 3 and 6, lists will be made with ratios of 1:2 for placebo to active. These lists will be created for each strata.

When an “active” slot is pulled a “coin flip” approach will be taken in order to select which active dose is selected. These probabilities will change monthly. These RAR probabilities will adjust for the 3 strata factors. The balancing approach for the RAR is described in Section “Balance Weighting.”

Part 2:

In Part 2 of the trial there will be only one target active arm and placebo. During this part randomisation is done using random blocks of 2, 4, or 6 with equal randomisation (1:1) within each site, for each of the 8 strata.

The individual lists created at each stage of randomisation will be discontinued at the end of its stage, with the new lists within site being used.

Balance Weighting

The adaptive randomisation algorithm creates a vector of probabilities for the four active treatment arms. This is a "global" probability for the arms, meaning it is over all strata. This section describes the balancing of the adaptive randomisation across the different strata to maintain, as well as possible, the balance of the 8 strata within active treatment arms, while achieving the needed adaptive randomisation.

The 8 strata used to balance the randomisation are based on the two-way classification of mechanical ventilation (MV) (Yes/No), Creatinine (Low/High), and norepinephrine/noradrenaline (NA) use (Low/High). We present the method for modifying the global adaptive randomisation probabilities to create different randomisation probabilities for each strata that will honor the goals of the response adaptive randomisation as well as the goal of balancing the strata in the treatment arms.

The outline of the approach at any interim point of the trial is:

1 The response adaptive randomisation probabilities are provided from the efficacy analysis. Label these π_1, π_2, π_3, and π_4, for the four active doses.

2 The odds-ratio of a stratum, for each arm, relative to all other arms, is calculated for the *previously* randomised patients.

3 Within each stratum the response adaptive randomisations are modified by the odds-ratios from the previous randomisations to create new odds for each arm that are strata specific.

4 The probabilities within each strata are normalized and used for randomisation.

The details of the balancing algorithm are presented below and an example data set is presented.

We label the 8 strata as s=1,…,8. For the previous randomisations, let N_{sa} be the number of stratum s randomised to arm a. The marginals of each stratum and/or arm are labeled with a + symbol.

The odds-ratio (modified by adding 0.5 to each cell for handling 0 counts) of the previous patients being randomised to arm *a*, within stratum *s* is labeled OR_{sa}:

$$OR_{sa} = \frac{\left(N_{sa} \middle/ \left(N_{+a} - N_{sa} \right) \right)}{\left(\left(N_{s+} - N_{sa} \right) \middle/ \left(N_{++} - N_{+a} - N_{s+} + N_{sa} \right) \right)}.$$

The global odds of randomising to each arm ($\pi_a/1-\pi_a$) for new patients is modified by the previous odds-ratio of randomisation to create the new randomisation probabilities by stratum for balancing future patients. The modified odds of randomising to arm *a*, within stratum *s* is

$$\left(\frac{\pi_a}{1-\pi_a} \right) \left(\frac{1}{OR_{sa}} \right).$$

These odds create unique probabilities for each arm, within each stratum.

$$\frac{\left(\frac{\pi_a}{1-\pi_a} \right) \left(\frac{1}{OR_{sa}} \right)}{1 + \left(\frac{\pi_a}{1-\pi_a} \right) \left(\frac{1}{OR_{sa}} \right)}$$

These are normalized (across a stratum to sum to 1) to form a probability distribution for arms within a stratum:

$$\Pr(\text{Arm a in Stratum s}) = \frac{\frac{\left(\frac{\pi_a}{1-\pi_a} \right) \left(\frac{1}{OR_{sa}} \right)}{1 + \left(\frac{\pi_a}{1-\pi_a} \right) \left(\frac{1}{OR_{sa}} \right)}}{\sum_{b=1}^{4} \frac{\left(\frac{\pi_b}{1-\pi_b} \right) \left(\frac{1}{OR_{sb}} \right)}{1 + \left(\frac{\pi_b}{1-\pi_b} \right) \left(\frac{1}{OR_{sb}} \right)}}$$

As an illustrative example, assume the high dose has not been opened (and the new randomisation probability π_4 is 0). Example of previous randomisations are presented in Table 2:

Table 2: Example of randomisations to each arm within each stratum.

Strata	Arms			Total
	a=1	a=2	a=3	
1	5	5	5	15
2	9	1	5	15
3	2	2	2	6
4	1	0	0	1
5	15	18	20	53
6	2	8	8	18
7	7	8	9	24
8	3	4	1	8
Total	44	46	50	140

The odds-ratios of each arm, within each stratum, OR_{sa}, are presented in Table 3:

Table 3: The odds-ratio of arm *a*, within each stratum, *s*, with the 0.5 factor added to each cell

Strata	Arms		
	a=1	a=2	a=3
1	1.09	1.02	0.90
2	3.42	0.18	0.90
3	1.09	1.02	0.90
4	3.32	0.51	0.45
5	0.79	1.08	1.15
6	0.28	1.69	1.48
7	0.88	1.02	1.09
8	1.28	1.92	0.32

Assuming the global response adaptive randomisation probabilities are π_1=0.33, π_2=0.50, and π_3=0.17, the modified randomisation probabilities for each arm within each stratum are presented in Table 4.

Table 4: The modified randomisation probabilities for each stratum

Strata	Arms		
	a=1	a=2	a=3
1	0.31	0.50	0.19
2	0.11	0.73	0.16
3	0.31	0.50	0.19
4	0.12	0.60	0.28
5	0.38	0.47	0.15
6	0.56	0.33	0.11
7	0.35	0.49	0.16
8	0.27	0.34	0.39

Appendix 8 Bias on Treatment Estimate for P&VFDs

The treatment estimates for the primary endpoint will be based on a comparison between all patients on all selepressin arms from both parts of the trial (pooled together and treated as a single arm), and all patients on the placebo arm from both parts of the trial.

Additionally, the primary analysis will be repeated for the selected arm only, i.e. comparing all patients on the selected arm (from part 1) from both parts of the trial (pooled), to all patients on the placebo arm from both parts of the trial.
To address the issue of a potential bias on the treatment estimates a variety of different scenarios were simulated (Table 5). The results are based on 3000 simulated trials of scenarios where all arms are equally effective and where the mortality benefits and P&VFD benefits for survivors correspond. For each simulated trial, the treatment effect estimates are simple classical point estimates from the raw patient data, e.g. the mortality treatment effect is estimated by the raw mortality rate in the placebo group, minus the raw mortality rate in the active group.

Note that when investigating bias, it is not appropriate to restrict attention to simulated trials that are successful. Estimates from that approach will yield results that are larger than the simulated truths, and this is true for simple non-adaptive designs as well. In this trial the design is modified by removing all futility rules so that all simulated trials select an arm for Part 2 and enroll all 1800 patients. Some of these simulated trials reach 800 patients with the placebo arm outperforming all active arms; in these Part 2 is begun with the least badly performing active arm.

The table shows the true underlying benefit, the estimated benefits for the primary analysis (all selepressin arms from both parts of the trial compared to all patients on placebo from both parts of the trial), and the estimated benefits for the selected arm only (all patients on the selected arm from both parts of the trial compared to all patients on placebo from both parts of the trial).

The placebo arm was assumed to have a 25% mortality rate, a mean of 24 P&VFDs for survivors, and an overall mean of 18 P&VFDs (survivors and non-survivors).

If all active arms are included in the estimates (to correspond with the use of all active arms in the final van Elteren analysis as in the two middle columns), the estimates are unbiased and any differences from the truth are due to Monte Carlo variation. If only the selected arm is included in the estimates, there is some bias due to the fact that the selected arm must have performed relatively well in Part 1, the fact that Part 1 data are included in the estimates, and this bias is not completely neutralized by the introduction of a large sample of unbiased data in Part 2. The estimates of benefit in P&VFDs for survivors are nearly unbiased, while the relatively noisier mortality estimates have biases that can be on the same scale as the treatment effect. The largest bias is for the small mortality effect of 0.5%, where the design estimates an average effect of 0.77%, which is too small of a benefit to lead to a successful trial without a substantial effect on P&VFDs.

Table 5: Simulated treatment estimate bias for P&VFDs

True benefit		**Average estimated benefit (all selepressin arm)**		**Average estimated benefit (selected arm only)**	
Mortality	P&VFDs (survivors)	Mortality	P&VFDs (survivors)	Mortality	P&VFDs (survivors)
0%	0 days	-0.03%	0.01 days	0.21%	0.06 days
0.5%	0.5 days	0.56%	0.51 days	0.77%	0.56 days
1.0%	1.0 days	0.98%	0.98 days	1.17%	1.05 days
1.5%	1.5 days	1.45%	1.46 days	1.63%	1.54 days
2.0%	2.0 days	1.92%	1.95 days	2.05%	2.03 days
3.0%	3.0 days	2.97%	2.97 days	3.10%	3.03 days

In both scenarios, treatment estimates are either unbiased or the treatment bias is negligible, and hence treatment estimates will not be adjusted for treatment bias.

Statistical Analysis Plan Addendum

Title of trial: A Double-blind, Randomised, Placebo-controlled, Phase 2b/3 Adaptive Clinical Trial Investigating the Efficacy and Safety of Selepressin as Treatment for Patients with Vasopressor-dependent Septic Shock
NCT number: NCT02508649
Sponsor trial code: 000133
Date: 20 Nov 2017

STATISTICAL ANALYSIS PLAN

ADDENDUM

A Double-blind, Randomised, Placebo-controlled Phase 2b/3 Adaptive Clinical Trial Investigating the Efficacy and Safety of Selepressin as Treatment for Patients with Vasopressor-dependent Septic Shock

SEPSIS-ACT

Selepressin Evaluation Programme for Sepsis-Induced Shock - Adaptive Clinical Trial

000133

Investigational Product:	Selepressin; concentrate for solution for infusion Placebo; sterile 0.9% sodium chloride solution
Indication:	Vasopressor-dependent Septic Shock
Phase:	2b/3
Author:	[redacted]
Date of issue:	November 20th - 2017
Version:	Final – Ver. 1.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes
1	Nov 20-2017	Original SAP addendum	Not applicable

Signed agreement on Statistical Analysis Plan Addendum

The original analysis plan addendum was reviewed by,

- [redacted] Global Biometrics, Ferring Pharmaceuticals A/S
- [redacted] Global Clinical Operations, CDS, Medical Writing & Selepressin, , Ferring Pharmaceuticals A/S
- [redacted] Medical Writing, Ferring Pharmaceuticals A/S
- [redacted] Selepressin, Ferring Pharmaceuticals A/S
- [redacted] Selepressin, Ferring Pharmaceuticals A/S

And reviewed and approved (signed electronically) by,

- [redacted] Biometrics, Ferring Pharmaceuticals A/S

TABLE OF CONTENTS

1 Introduction

This document describes the statistical analyses for selepressin (FE202158) 000133 planned after the trial was stopped for futility, but before the blind was broken. It is an addendum to version 9.0 of the statistical analysis plan.

1.1 Abbreviations

Abbreviations	Meaning of abbreviations in document
IMP	Investigational medicinal product
P&VFDs	Vasopressor- and mechanical ventilator-free days

2 Post-futility / Pre-unblinding Planned Analyses

2.1 Primary Endpoint

To explore whether there was a 'learning curve' effect of administering the drug, the primary analysis will be conducted without the first 3 patients at each site.

Due to trial logistics (e.g. preparation time for pharmacy etc.) or safety concerns, some patients may receive norepinephrine instead of the investigational medicinal product (IMP) during parts of the treatment period where they have a vasopressor need. To try to mimic the 'near perfect' conditions, where IMP is used as the primary vasopressor, the primary analysis will be conducted on a subset of the per-protocol analysis set in which patients who did not have norepinephrine weaned before IMP (unless due to death), or patients in which IMP was not restarted within 6 hours upon restart of norepinephrine, will be excluded.

To evaluate the new septic shock criteria (1), the primary endpoint will be evaluated for patients with lactate <2 mmol/L at baseline (sepsis) and patients with lactate ≥2 mmol/L (septic shock), respectively.

The VASST trial identified a potential survival benefit using vasopressin as compared to norepinephrine in patients with less severe septic shock defined as a baseline norepinephrine dose of less than 15 µg/min. Similarly, to assess whether there was a potential benefit of selepressin in vasopressor- and mechanical ventilator-free days (P&VFDs) in patients with less severe septic shock, the treatment effect of the primary endpoint will be presented graphically, in order to visually inspect whether the average treatment effect is distributed evenly across the baseline norepinephrine levels (≤15 µg/min, >15 µg/min).

To further assess if baseline norepinephrine impacts the treatment effect, the same analysis will be repeated for baseline norepinephrine levels (≤0.1µg/kg/min,]0.1; 0.20 µg/kg/min],]0.2; 0.30 µg/kg/min],]0.3; 0.45µg/kg/min],]0.45; 0.60µg/kg/min], >0.60µg/kg/min).

To assess if there is a treatment duration effect, patients will be stratified by cumulative duration of IMP infusion (<1 day, [1; 3 days[, [3; 7 days[, ≥7 days) and the treatment effect of the primary endpoint will be presented graphically, in order to visually inspect whether the average treatment effect is distributed evenly across the duration of treatment with IMP.

2.2 Organ Dysfunction

2.2.1 Mechanical Ventilator-free Days up to Day 30

To assess if there is an effect of baseline lung function on the potential treatment effect of selepressin on lung function, the treatment effect of mechanical ventilator-free days will be presented graphically, in order to visually inspect whether the average treatment effect is distributed evenly across baseline PaO_2/F_iO_2 levels (<100mmHg, [100; 200 mmHg[, [200; 300 mmHg[,≥300 mmHg).

2.2.2 Duration of Mechanical Ventilation up to Day 30

The graphical presentation in section 2.2.1 will be repeated for duration of mechanical ventilation.

2.2.3 Renal Replacement Therapy (RRT)-free Days up to Day 30 (excluding patients on RRT for chronic renal failure at time of randomisation)

RRT-free days will also be evaluated using the new septic shock definition (1) as for the primary endpoint.

2.3 Morbidity/Mortality

2.3.1 Intensive Care Unit (ICU)-free Days up to Day 30

ICU-free days will also be evaluated using the new septic shock definition as for the primary endpoint.

2.3.2 All-cause Mortality (Defined as the Fraction of Patients That Have Died, Regardless of Cause, by the end of Day 30, Day 90, and Day 180)

Mortality will also be evaluated using both the new septic shock definition, and the baseline norepinephrine cut-offs as for the primary endpoint.

2.4 Daily and Cumulative Urinary Output Until ICU Discharge (for a Maximum of 7 Days)

Urinary output and cumulative urinary output analyses will be repeated for the two components that make up total output volume; spontaneous urine output and output fluids collected from drainages, suction devices and respiratory fluid loss estimation (if transpiration was performed).

2.5 Pulmonary Function (PaO2/FiO2) (in a subset of patients)

PaO_2/F_iO_2 up to and incl. Day 7 will be evaluated as an area under the curve (AUC).
Linear interpolation will be used to derive the AUC (taking into account that not all patients will have measurements taken between 24 hours post baseline and Day 2). Patients dead or withdrawn will be set to missing. The cumulative PaO_2/F_iO_2 will be compared between treatment arms using a repeated measures ANCOVA model with baseline PaO_2/F_iO_2 as covariate, treatment, time and treatment by time interaction as factors, baseline PaO_2/F_iO_2 by time interaction, and patient as the experimental unit. Estimated treatment differences (to placebo) along with a 95% confidence interval will be presented.

2.6 Arterial Blood Gases and Acid/Base Status (PaO2, PaCO2, pH, SaO2, Bicarbonate, Base Excess), Lactate and Oxygen Saturation in Vena Cava Superior (ScvO2)

Lactate up to and incl. Day 7 will be evaluated as for PaO_2/F_iO_2 in section 2.5 above.

2.7 Time to Out of Shock

A Kaplan-Meier analysis will be performed for time to out of shock, defined as the first time where the patient has no vasopressor for at least 24 hours. Patients who die or are withdrawn while still on vasopressors and before first 24 hour vasopressor-free episode will be censored. The same applies for patients prematurely discontinued from IMP.

2.8 Time to Out of Mechanical Ventilation

Time to out of mechanical ventilation will be defined and analyzed similarly as for time to out of shock in section 2.7 above.

2.9 Proportion of Patients maintaining Target MAP while on IMP

The proportion of patients maintaining target MAP while on IMP will be compared between treatment groups using a logistic regression model.

2.10 Episodes of Hypotension

The proportion of patients having one or more episodes of hypotension will be compared between treatment groups using a logistic regression model.

The cumulative duration of periods of hypotension will (ideally) be compared between treatment groups using a zero inflated negative binomial model. In case model assumptions do not hold, a permutation test will be used.

3 References

1 Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016. 315(8):801-810.