Statistical Analysis Plan for: The EFFORT Trial

1 Administrative Information

1.1 SAP summary table

TRIAL FULL TITLE	The Effect of Higher Protein Dosing in Critically Ill Patients: A Multicenter Registry-based Randomized Trial: The EFFORT Trial
TRIAL REGISTRATION	https://clinicaltrials.gov/ct2/show/NCT03160547
PROTOCOL PUBLICATION	*The EFFORT Trial. JPEN J Parenter Enteral Nutr. 2019 Mar;43(3):326-334.*
CURRENT PROTOCOL DATE	Oct 27, 2017
TRIAL PRINCIPAL INVESTIGATOR	Daren K. Heyland
TRIAL SENIOR STATISTICIAN	Andrew G. Day
TRIAL COORDINATOR	Jen Korol
STATISTICIAN(S) PERFORMING ANALYSIS	Xuran Jiang and Andrew G. Day
SAP AUTHOR(s)	Andrew G. Day, Xuran Jiang, Jen Korol and Daren Heyland
SAP DATE	March 9th 2022
SAP STATUS	First finalized version (V1)
SAP REVISION HISTORY	None this is V1
STATUS OF TRIAL AT TIME OF SAP FINALIZATION OF V1	Enrollment completed. Blinded data cleaning essentially complete. No by arm outcome results generated yet.

1.2 Signatures

I have read and approve the enclosed SAP dated 2022-03-9 for EFFORT trial.

Senior Statistician & SAP Author

Name: Andrew G. Day

Signature: Andrew G. Day

Date: March 9, 2022

Statistician Performing Analysis (other than senior statistician):

Name: Xuran Jiang

Signature: Xuran Jiang

Date: 10Mar2022

Trial Co-ordinator

Name: Jennifer Korol

Signature: JKorol

Date: March 10, 2022

Principal Investigator

Name: Daren K. Heyland

Signature:

Date: Mar 9/2022

1.3 Purpose, usage, and target audience of this document

This document provides a detail description of the analysis plan for the EFFORT trial. This document is meant to be used in conjunction with the study protocol. This document does not subsume the protocol, but several elements of the protocol, such as the sample size justification are reproduced herein for completeness. This document has the following purposes:

1. Provides a written agreement between the principal investigator, study co-ordinator, lead study statistician and data analysts regarding exactly what analysis will be performed.
2. Provides a record of the analysis plan specified prior to examining any outcomes by arm.
3. Provides clear specifications for the analyst(s) performing the data filtering/transformation, variable derivations, statistical analyses and report generation.

This document follows the guidance published in JAMA by Gamble et al (2017) and referenced at https://www.equator-network.org/reporting-guidelines/guidelines-for-the-content-of-statistical-analysis-plans-in-clinical-trials/ (1) The SAP checklist is completed in Appendix B.

1.4 SAP contributors and signatories

Andrew Day drafted the SAP, Xuran Jiang contributed details regarding the definition of several outcomes, Jennifer Korol added details regarding the trial operation and data management, and Daren Heyland helped interpret the protocol and prioritize outcomes, analyses, and validation. All authors as well as Shawna Froese provided critical review and editing to all parts of the SAP. The finalized version of the SAP was approved and signed off by all authors.

Table of contents

2 Introduction to Study

2.1 Background and rationale

Clearly definitive proof from prospective RCTs evaluating different levels of protein intake in nutritionally high-risk patients is lacking. Moreover, based on the arguments for and against higher dose of protein administration, there is clinical equipoise or uncertainty about the optimal dose of protein in ICU patients. Hence, we believe a trial, such as the EFFORT trial, that evaluates the effect of high protein intake in high nutritional risk critically ill patients is warranted. See study protocol for full background and rational.

2.2 Primary Research Question

In critically ill patients with nutrition 'risk factors', what is the effect of prescribing a higher dose (⩾2.2 grams/kg/day) of protein/amino acid administration compared to a usual dose prescribed ⩽1.2 gram/kg/day on (1) time to discharge alive TTDA from hospital and (2) 60-day mortality.

2.3 Study hypotheses

Compared to receiving usual dose of protein/amino acids, the administration of a higher dose of protein/amino acids (a consequence of having a higher prescription) to nutritionally high-risk critically ill patients will be associated with a quicker rate of recovery and improved survival.

3 Study Methods

3.1 Trial design

A multicentre, pragmatic, volunteer-driven, registry-based, randomized, open-label, clinical trial with patients randomized from 82 sites in from the following countries.

Country	Number of sites
Argentina	2
Australia	1
Brazil	3
Canada	10
Greece	1
Hong Kong	1
India	1
Iran	1
Japan	2
Malaysia	1
Mexico	4
Panama	2
Puerto Rico	1

Saudi Arabia	1
United Kingdom of Great Britain and Northern Ireland	31
United States of America	20

3.2 Modification to trial design from initial protocol

Due to COVID-19 pandemic, our original enrollment goals become unfeasible. Consequently, the study steering committee made the decision on September 15, 2021 to end study enrollment at 1, 200 patients. This made the effect size needed to achieve adequate power for the primary outcome of 60-day mortality unrealistic. However, under realistic assumptions, there may still be adequate power to detect an effect on the secondary outcome of TTDA (see section 3.4). For this reason, the primary and secondary outcomes were swapped.

3.3 Randomization

Patients were screened, evaluated, and randomized within 96 hours of admission to the ICU. Depending on the local REB consent was either waived or obtained from a legal representative. The site representative logged into the web-based central REDCAP randomization tool to confirm eligibility and receive the treatment assignment.

The randomization system used a computer-generated randomization schedule created by the trial statistician who was involved in recruitment and unaware of which stratum numbers mapped to which sites. Randomization was 1:1 to either high dose protein prescription or usual dose using permuted blocks of random size within strata (site). The random block sizes were either 2, 4 or 8 but this was not disclosed until the study enrollment was complete. Given the large pragmatic nature of the trial, site was the only stratification factor.

3.4 Sample Size Considerations

The original planed sample size of 2,000 patients per arm was chosen to provide 80% power at a two-sided alpha=0.05 to detect a 60-day mortality reduction from 30% to 26%.

With the new sample size of 600 patients per arm we would require a mortality reduction from 30% to 22.9% (a 24% relative risk reduction) to achieve 80% power. Such a large effect may be unrealistic and is larger than what we'd consider minimally clinically import. Therefore, we switched the primary and secondary outcome, so now the primary outcome is time to discharge alive (TTDA) from hospital.

At the time the steering committee reconsidered the sample size, we estimated the pooled 60-day mortality rate to be 32% and the median TTDA to be 21 days based on the 228 patents finalized at that time. This analysis did not require unblinding of the treatment arms since no by-arm analysis of outcomes were performed. Based on 1000 simulations using the aforementioned assumptions and assuming that among survivors TTDA followed an exponential distribution, we estimated that a sample size of 600 patients per arm would achieve e 83% power at a two-sided alpha=0.05 if there was a 15% relative risk reduction in hospital mortality from 34.6% to 29.5% combined with a 20% increase in the hazard rate of TTDA among hospital survivors. With a 10% relative risk reduction in mortality from

33.7% to 30.3% we would need the hazard of TTDA to be increased by 30% to achieve 85% power. Detail sample size considerations including the estimated power across various effect sizes are provided in appendix A with Table 2 providing results for a sample size of 600 patients per arm.

3.5 Framework

This is a confirmatory (i.e. hypothesis testing) two-sided superiority RCT comparing the efficacy higher versus usual protein/amino acid.

3.6 Interim analyses

The original protocol which planned to enroll 4000 patients, had a planned interim analysis at 2000 patients. However, no interim analysis was performed because this study was stopped after 1, 200 patients were enrolled.

3.7 Timing of final analysis

All outcomes will be analyzed by arm once all data is collected, data cleaning is complete and after this statistical analysis plan is finalized and posted at https://clinicaltrials.gov/ct2/show/NCT03160547.

3.8 Timing of outcome assessments

All outcomes were assessed while in hospital up to a maximum of 60 days post ICU admission except daily nutrition was assessed only for the first 12 days in the ICU and daily protein intake was reported up to 28 days post ICU admission.

3.9 Confidence intervals and P-values

95% confidence will be presented for selected key outcomes. P-values will be two-sided without adjustment for multiplicity. We will use the traditional two-sided $p \le 0.05$ to indicate statistical significance for the primary outcome. We will consider the multiplicity of comparisons when interpreting the significance of the additional outcomes and may even avoid reporting p-values of additional outcomes depending on journal guidelines.

3.10 Adherence and protocol deviations

The following are the key metrics that we reported on throughout the study and will be reported overall (not by site) in the final statistical report:

1. **Intervention**

- timing from ICU admission to randomization (hours)
- the number of nutrition evaluable days post randomization
- Average (range) of protein prescribed per day / Kg since randomization
- Average (range) of calories prescribed per day / Kg since randomization
- compliance with study intervention (gm of protein/kg, % of amount prescribed)
- other protocol violations:
  numbers in High dose group not reaching the 80% of prescribed protein
  numbers in Usual dose reaching high doses of protein (>100% of prescription)
  Patients received $\geq$ 110% of energy targets (proportions) since randomization

2. **Nutritional indices**

- Time from ICU admit to start of EN (hours)
- Days on EN
- Time from ICU admit to start of PN (hours)
- Days on PN
- Type of nutrition
    - Both EN and PN
    - EN only
    - PN only
    - None

- **Nutrition adequacy (protein and energy)**

  Nutritional adequacy will only be calculated for evaluable days. Evaluable days include the day after ICU admission up to the earliest of the day prior to ICU discharge (or death) or the first day patients began exclusive permanent oral feeding for the remainder of the first 12 ICU days. Nutritional adequacy will be calculated daily for each of the evaluable days as the total calories and protein (as separate outcomes) received from all collected sources divided by the corresponding baseline prescriptions and multiplied by 100. Evaluable days without any nutrition support received will be counted as 0 adequacy for both energy and protein.

  Total energy will be calculated as the sum of all calories from EN, PN propofol (estimated as 1.1kcal/mL), supplemental protein, non-protein modular supplements and amino acids. Total protein will be calculated as the total of grams of protein from EN, PN, EN supplements and PN amino acids.

  We will create curves depicting the mean daily energy and protein adequacy from ICU admission to day 12 as well as from randomization to ICU day 12 (which will generally be from 8-12 days after randomization). We will also use descriptive statistics to summarize the energy and protein adequacies averaged over patients' evaluable ICU days as well as averaged over evaluable days excluding days prior to randomization.

- Use of motility agents in patients with averaged caloric adequacy of <80% during their evaluable days.
- average grams/kg/day of protein by each of its constituents (EN, PN, EN supplements and PN amino acids) and in total.

3. **Completeness of data capture**
    - Completeness of Case Report Forms (100%) – does charet have any missing data.
    - missing data for baseline demographics
    - missing ICU LOS
    - missing Hospital LOS
    - missing mechanical ventilation days
    - missing date/time to discharge alive from hospital
    - missing in hospital mortality and 60-day mortality

3.11 Analysis populations

In accordance with the intent-to-treat principle, the primary analysis will include all patients in the arm to which they were randomized regardless of study compliance. However, a priori, we planned to exclude patients who were randomized in error and randomized patients who did not receive the study intervention. In addition, there are some patients that were randomized and due to lack of compliance from the site, we have insufficient information to know whether they received the study intervention; thus, our analysis will be a modified ITT. Patients with adequate nutrition information but missing outcome data were included. Also, we plan an efficacy analysis in which we will only include patients treated as per protocol. That is, they remained on artificial nutrition for at least 4 calendar days and achieved at least 80% of their prescription in the high dose group, received no more than 1.2 grams/kg/day in the usual dose group and received not more than 110% of energy requirements in both groups.

4 Trial population

4.1 Eligibility criteria:

Published at https://clinicaltrials.gov/ct2/show/NCT03160547

4.2 Screening, recruitment, patient flow/follow-up

A CONSORT style flow diagram will present the numbers of patients screened and all reasons excluded prior to randomization (2). This figure will also present the number randomized to each arm and then will work down to the numbers included in the assessment of the primary and secondary outcomes with counts of all post-randomization exclusion reasons. In addition, a table will provide the study ID error code (if applicable) and description of the reason for each randomized patient being excluded from the evaluation of the primary or secondary outcome.

4.3 Baseline characteristics

Baseline characteristics will be described by arm and overall using descriptive statistics only. Categorical variables will be described as counts (%). Continuous variables will be described as mean±SD (min to max) and/or median [Q1 to Q3].

The following baseline patient characteristics will be described:

Age, sex, BMI, admission category, primary ICU diagnosis, Charlson comorbidity index, baseline SOFA Score, APACHE II score, mNUTRUC score, baseline assessment of malnutrition, Clinical Frailty Scale, Sarcopenia, and geographic region (Canada/Australia, USA, Other Americas, United Kingdom and Northern Ireland, Asia).

5 Analysis

5.1 Outcome definitions

See section 3.2 which describes the modification of the sample size and switching of the primary and secondary outcomes due enrollment ceasing during the COVID-19 pandemic.

5.1.1 Primary outcome:

Time to discharge alive from hospital [time frame 60 days from ICU admission].

5.1.2 Secondary outcomes:

60-day mortality. All patients discharged alive from hospital were considered alive at day 60 unless there was information to suggest otherwise.

5.1.3 Other registered outcomes:

1. Nutritional Adequacy [Time Frame: 12 days for calories for and 28 days for protein]
2. Index hospital mortality [Time Frame: up to 60 days post ICU admission]
3. Readmission to ICU and Hospital [Time Frame: up to 60 days post ICU admission] readmit to ICU or hospital defined as ≥ 48 hrs from last admission
4. Duration of Mechanical Ventilation [Time Frame: up to 60 days post ICU admission and considers ventilation continuing if re-ventilated with 48 hours of extubation.
5. ICU length of stay [Time Frame: up to 60 days post ICU admission]
6. Hospital length of stay among 60-day survivors [Time Frame: up to 60 days post ICU admission]

5.1.4 Additional unregistered outcomes:

POD+death free days after randomization and up 30 and 50 days after randomization as well as precent with POD+death at day 30 and 50 after randomization.

5.1.5 Serious adverse events:

Not collected for this pragmatic registration study.

5.1.6 Economic evaluation:

None planned

5.2 Analysis Methods

5.2.1 Primary outcome

The primary outcome is time to discharge alive (TTDA) from hospital up to 60 days after ICU admission. Death will be considered a competing risk precluding live discharge. TTDA or death will be counted from time of randomization even though follow-up only goes to 60-days after ICU admission, thus patients still in hospital at ICU day 60 may will be censored up to 96 hours before day 60 depending on how long after ICU admission they were randomized. For patients who are readmitted to the index hospital within 60-days (that is for all of those who have a 2nd or 3rd hospitalization collected), we will use the discharge information from the final recorded hospitalization. Patients known to have died within 60-days of ICU admission will be considered decedents never discharged alive regardless of prior discharge. Patients who did not die within 60 days of ICU admission and who are not still in hospital at 60 days after ICU admission will be considered to have been discharge alive at their last hospital discharge date unless their discharge location is: ICU in another ward, Ward in another hospital, or “other” with a description including the word “hospice”. For these 3 final discharge locations we will censor patient at the time of discharge. If patients were lost to follow-up prior to death or hospital discharge, then we will censor them at the last date we know they were in the hospital. Patients who withdrew consent for prospective data collection while in the hospital will be censored at date of consent withdrawal. The cumulative incidence function (CIF) curves will be displayed by arm and the difference in the CIF between arm will be tested by the Gray test as implemented in the SAS LIFETEST procedure (3). Gray’s test is essentially a log-rank test where decedents are censored after the end of follow-up (i.e. after 90 days). The median and quartiles of time to live discharge from the CIF will be reported by arm. The between arm difference will be summarized by the unadjusted subdistribution hazard ratio as estimated by the Cox proportional hazards model (4). A sensitivity analysis will use a shared frailty model to incorporate ICU as a random effect (5). This shared frailty analysis will censor decedents at 61 days (after end of follow-up) which will mimic the Fine and Gray approach and provide a subdistribution hazard ratio while controlling for ICU. The subdistribution hazard ratios will be reported with 95% confidence intervals as estimated by PROC PHREG in SAS.

5.2.2 Secondary outcome

We will compare 60-day mortality between arms using Pearson’s chi-square test for two independent proportions. Patients discharged alive from hospital with unknown survival status at 60 days will be assumed to be alive. Unadjusted relative risks and hazard ratios will be presented with 95% CIs. A secondary analysis will employ the generalized mixed effects model with a random site effect using a log link and binomial distribution (i.e. the mixed log-binomial model) to provide adjusted estimates of relative risk with 95% CIs. This will provide a within site interpretation of effect, will allow us to explore between site heterogeneity and will meet regulatory guidance suggesting that site be incorporated in a sensitivity analysis if it is not used for the primary analysis. (6-8) However, if this model does not coverage, we will use a logit link (i.e. mixed logistic regression) which will provide estimates of odds ratios.

5.2.3 Other registered outcomes

Nutrition adequacy will be reported with the adherence measures as described in section 3.1

Index hospital mortality, readmission to ICU and readmission to hospital will be reported by arm as counts and percentages with differences tested by Chi-Squared test or Fisher’s exact test if there are less than 20 people with the given event across both arms.

For durations of mechanical ventilation, ICU stay and hospital stay we will report the median [Q1, Q3] time after randomization by arm summed over all recorded hospital stays. These will also be reported in the subset of patients who did not have a death reported within 60 days.

5.2.4 Additional unregistered outcomes

The mean (SD) of modified POD free days by 30 and 50 days of randomization will be reported by arm. Also, the count and proportion of patients POD+death free at day 30 and day 50 will be reported by arm. The POD free days by day 30 and 50 will be compared between arms by the Wilcoxon Rank-sum test.

We will start counting POD free days on the day after randomization. Patients who are known to have died (i.e. died within 60 days of ICU admission or 56-60 days of randomization) will be considered to have 0 PODS free days.

A patient will be considered POD free on a given day if they are not mechanically ventilated, on vasopressors or on renal replacement therapy according to the following definitions:

(1) **Mechanical ventilation**: if any part of the calendar day is on or between the start and stop date of any invasive mechanical ventilation period, or the patient restarted invasive mechanical ventilation within 48 hours of the current day, then then the day is not a free day.

(2) **Vasopressor therapy**: days with more than 2 hours of any dose of norepinephrine, epinephrine, vasopressin, Dobutamine, Milrinone or Levosimendan and >5 ug/kg/min of dopamine, or > 50 ug/minute of phenylephrine, will not be considered free days. The 48-hour rule does not apply to vasopressor therapy or renal replacement therapy.

(3) **Renal replacement therapy**: if any part of the calendar of the calendar day is on or between the start and stop date of any renal replacement therapy period then the day is not a free day.

5.3 Adjustment for covariates

For the primary and secondary outcomes, we include a sensitivity analysis that control for ICU as a random effect. We will also perform a sensitivity analysis of the primary and secondary outcomes adjusting for age, APACHE II score, mNUTRIC score, clinical frailty score, sarcopenia (SARC-F), admission type and geographic region in addition to site as a random effect. For the primary time to hospital discharge alive outcome, this will use the shared frailty model describe in section 5.2.1 and for 60-day mortality this will use the generalized mixed effects model with a random ICU effect described in section 5.2.2.

5.4 Assumption checking

The proportional hazards (PH) assumption of the primary outcome (time to live hospital discharge) will be assessed visually based on the roughly parallel CIF curves and log-negative log survival vs. log of time. Violations of the PH assumption do not invalidate the tests, but complicate the interpretation of the hazard ratio. If there is an important violation of the PH proportion hazard assumption then, emphasis will be placed on the overall CIF curves and the median time to live discharge rather than the subdistribution hazard ratio.

The only assumption for the binary outcomes is that missing data is are missing completely at random (see section 5.6 for assessment of this) and that each observation is independent. We have a planned sensitivity analysis that will control for site as a random effect, and we will examine the primary and secondary outcome by geographical region with a test for effect modification by region by modelling a region by treatment arm interaction effect in the multivariable models.

5.5 Subgroup analyses

We will perform a pre-specified subgroup analysis based on baseline NUTRIC score (0-4 vs. 5-9). In addition, we plan to evaluate the treatment effect within subpopulations of our enrolled patients (age (based on median), sepsis, burns, trauma, acute kidney injury or use of RRT, severity of illness [median APACHE] and BMI>30), depending on the numbers of patients in each of these subgroups. Finally, we will consider the effect of each of the multiple nutrition risk factors, both individually and in combination on the magnitude of the treatment effect. The statistical significance of apparent effect modification will be assessed by testing a treatment by covariate interaction term using logistic regression for mortality and Cox PH model accounting for competing risk of death for time to discharge alive. Due to the increased risk of type I and type II error, subgroup specific inferences will be considered exploratory and hypothesis generating. Subgroup specific effects may be presented by forest plots. An additional post-hoc subgroup analysis will be performed by region (Canada/Australia, USA, Other Americas, United Kingdom and Northern Ireland, Asia).

5.6 Missing data

The number of missing (or conversely non-missing) values will be reported by arm for every outcome. For the primary outcome (time to live hospital discharge) we expect minimal loss to follow-up since this outcome is not followed beyond the index hospital admission. The reasons of all missing primary and secondary outcomes will be reported by arm. Patients lost to follow-up in hospital will be included but censored at the time of LTFU. For the primary outcome, we will perform an extreme case sensitivity analysis that first assumes all patients lost to follow-up (LTFU) in one arm were discharged alive at the time of LTFU and all patients in the other arm died at time of LTFU. If these two extremes yield similar conclusions, then no further missing data sensitivity analysis will be performed. If these two extremes yield differing conclusions, then we will perform a graphical pattern mixture tipping point approach demonstrating the treatment effect over the possible range of missing outcomes. (9, 10)

5.7 Additional analysis

The database generated from the EFFORT trial may be used for additional secondary analyses exploring questions other than assessing the efficacy of higher vs. usual protein/amino acid in critically ill patients in the ICU. Plans for these additional secondary analyses are to be determined and are not part of the primary EFFORT analysis or the corresponding primary publication.

5.8 Statistical software

The main analysis was performed using SAS 9.4 TS level 1M2 and SAS/STAT version 15.1 under Windows 7 Professional version 10.0.18362. The independent validation of selected items (see section 8.2) was performed using the same software and operating system except SAS 9.4 was level TS1M6.

6 Quality Assurance

6.1 Data quality

Data was entered into REDCap by local site personal. Each user with access to REDCap had a unique username and password. Access to REDCap was secure and an audit trial was maintained to keep track of the username, time, and values of all data entry and modification. A custom query module was used to implement extensive value, range, logical (including date sequence) data checks. Any violation of the pre-defined data checks triggered data queries that were tracked and required resolution (either correction or acceptance by central staff) prior to data being marked as finalized.

Quality assurance reports were run periodically throughout the trial to assess the completeness, timeliness, validity and quality of trial implementation and data capture by site. After a minimum of 2 patients in each group were finalized, sites were assessed for risk and were categorized high, medium or low risk. Issues were flagged with participating sites, and further follow-up was conducted as needed

High-priority items from 2 patients (one in each group) at each site were monitored via remote source verification once they had randomized 2 patients, except for those sites that didn't allow remote monitoring.

The REDCap database was downloaded and converted into a multi-table relational analytic SAS database. Some filtering, data transformation, and variable derivation was performed in SAS. Boxplots were generated for all continuous variables and outliers were queried; all outliers were either corrected or verified as correct.

6.2 Validation of SAS database and analysis

The study PI and study co-ordinary will sense check all results to make sure they are not highly suspicions and that all counts are consistent with the patient flow diagram.

A second statistician who did not perform the primary analysis will independently verify the patient flow counts and re-analyze the primary and secondary outcomes of: 1) Time to discharge alive from hospital and 2) 60-day survival.

7 References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-2343.
2. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG; Consolidated Standards of Reporting Trials Group. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010 Aug;63(8):e1-37.
3. Gray, R. J. (1988). "A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk." Annals of Statistics 16:1141–1154.
4. Fine, J. P., and Gray, R. J. (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk." Journal of the American Statistical Association 94:496–509.
5. Duchateau L, Janssen P, Lindsey P, et al. The shared frailty model and the power for heterogeneity tests in multicenter trials. Computational Statistics & Data Analysis 2002; 40: 603–620.
6. Committee for Proprietary Medicinal P. Points to consider on adjustment for baseline covariates. Statistics in Medicine. 2004;23:701-709.
7. Hubbard WK. ICH: Guidance on Statistical Principles for Clinical Trials (FDA Docket No. 97D-0174). Federal Register. 1998;63:49583-49598.
8. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? BMC Medical Research Methodology. Feb 2014;14.
9. Liublinska V and Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial.
10. Hollis S. A graphical sensitivity analysis for clinical trials with non-ignorable

8 Appendix A: Updated sample size considerations

The reminder of this appendix was prepared for the steering committee when it became clear that the original sample size of 4000 patients was unachievable because of near cessation of enrollment due to the COVID-19 pandemic.

With 2000 patients per arm as originally proposed, the EFFORT trial would achieve 80% power to detect a 4% absolute risk reduction from 30% to 26% (13.3% relative risk reduction) in 60-day mortality using a Chi-Squared test at a two-sided alpha=0.05.

The figure and tables below provides power estimates (at two-sided alpha=0.05) for mortality and time to discharge alive (TTDA) outcomes for various effect sizes and sample sizes. The arm specific mortality and TTDA among survivor estimates are based on the current (n=228 finalized) pooled 60-day mortality rates of 32% and median TTDA of 21 days assuming the TTDA among survivor followed an exponential distribution. All estimates are based on 1, 000 simulations. Note that a RR<1 and a TTDA hazard rate (HR)>1 are beneficial as they represent a reduced rate of mortality and an increased rate of TTDA.

With a total sample size of 1,000 (Table 1 and first panel of figure), a 25% mortality relative risk reduction (RRR) from 36.5 to 27.4, we achieve 87% power for testing for mortality, and power increases using TTDA if and only if the hazard ratio for TTDA among survivors is at least 10% higher in the intervention arm compared to the control arm. If the RRR in mortality is 15%, so that morality is reduced from 34.6% to 29.4% then we do not have adequate power to compare mortality, but we'd achieve 79% power for TTDA if the hazard ratio of TTDA was at least 20% higher in the intervention arm than the control arm. Generally, TTDA has more power than mortality alone if the hazard rate for TTDA among survivors is at least 10% higher in the intervention arm compared to the control arm. However, a smaller effect size in TTDA results in a loss of power compared to using mortality alone.

With a total sample size of n=2000 (table 4 and last panel in figure) using TTDA we are good for any TTDA beneficial HR if the mortality RR 0.8 or less (red and blue lines in last panel of figure). If RR of mortality is 0.85 (i.e. 15%RRR – green line in figure) then we're adequately powered if the intervention accelerates the daily change of discharge by 9% or more. If there is only a 10% RRR in mortality (brown line) , then we need the intervention to increase the daily live discharge speed by almost 20% to get above 80% power. If there is no difference in survival then TTDA will be underpowered even if we accelerate discharge by 30%.

With a total sample size of 1500 (table 3 and 3rd panel of figure), we'd achieve approximately 90% power if we had a 25% RRR in mortality with no benefit in TTDA, a 20% RRR in mortality with a 10% increase in the rate of TTDA, a 15% RRR in mortality with a 20% increase in the rate of TTDA, or a 10% RRR in mortality with a 30% increase in TTDA rate.

Figure 1: Power of Time to Discharge Alive

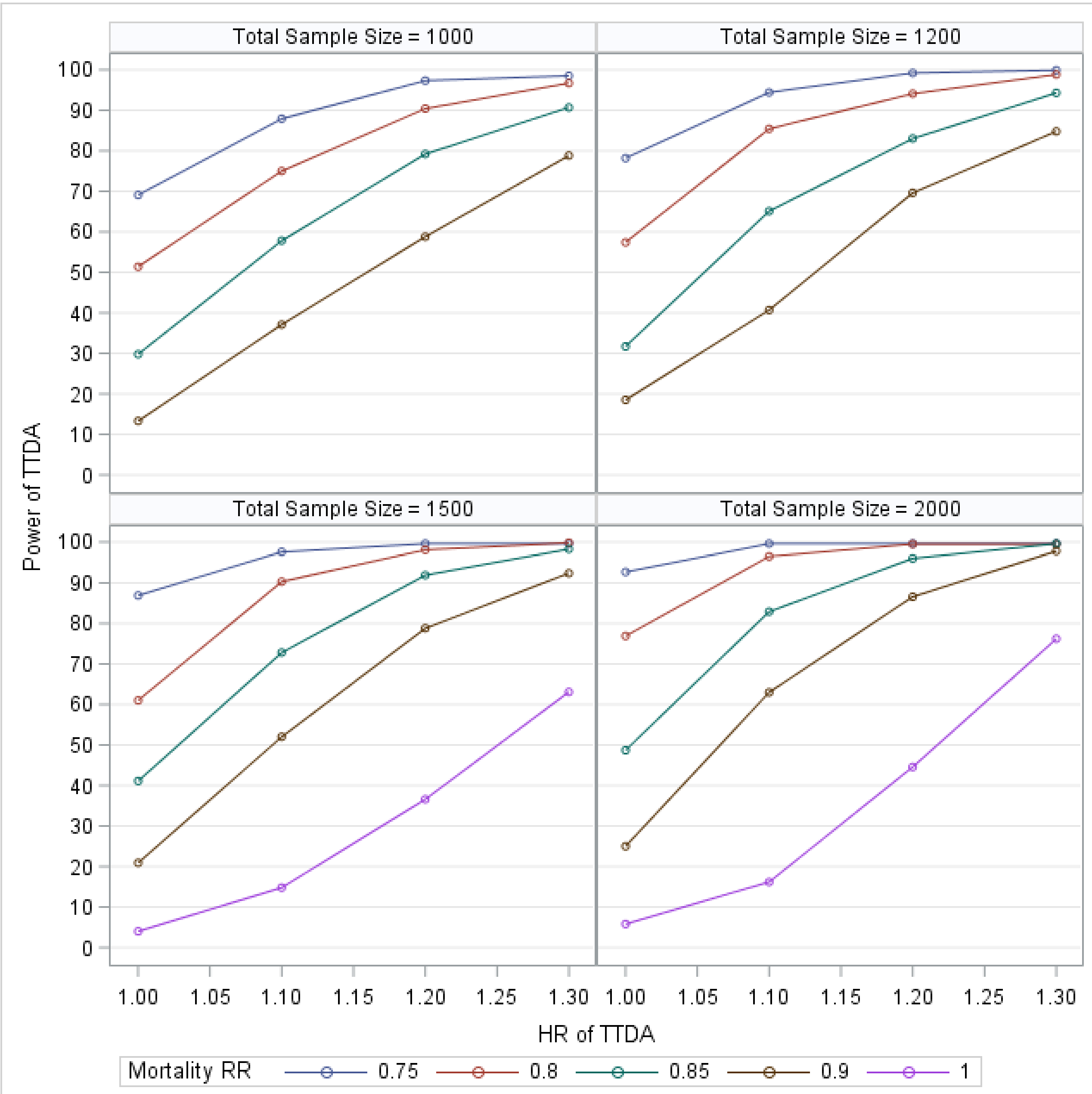

Table 1: Power Estimates with n=500 Per Arm

Assumed Effect Size		60-Day Mortality		Median TTDA Among Survivors		Power	
Mortality HR	TTDA HR	Control	Intervention	Control	Intervention	Mortality	TTDA
0.75	1	36.5	27.4	20.9	20.8	87%	69%
0.75	1.1	36.5	27.4	21.8	19.8	87%	88%
0.75	1.2	36.6	27.4	22.7	18.9	88%	97%
0.75	1.3	36.6	27.4	23.6	18.1	87%	99%
0.8	1	35.6	28.4	20.7	20.8	69%	51%
0.8	1.1	35.5	28.4	21.8	19.9	67%	75%
0.8	1.2	35.6	28.5	22.6	18.9	68%	90%
0.8	1.3	35.5	28.4	23.5	18.5	67%	97%
0.85	1	34.6	29.4	20.8	20.8	45%	30%
0.85	1.1	34.6	29.4	21.8	19.7	42%	58%
0.85	1.2	34.7	29.5	22.7	18.8	41%	79%
0.85	1.3	34.5	29.4	23.5	18	42%	91%
0.9	1	33.8	30.3	20.8	20.8	22%	13%
0.9	1.1	33.7	30.2	21.8	19.8	21%	37%
0.9	1.2	33.7	30.4	22.6	18.8	22%	59%
0.9	1.3	33.6	30.4	23.6	18.1	20%	79%

Green good power, yellow acceptable power and red unacceptable power. Assumes two-sided tests at alpha=0.05.

Table 2: Power Estimates with n=600 Per Arm

Assumed Effect Size		60-Day Mortality		Median TTDA Among Survivors		Power	
Mortality HR	TTDA HR	Control	Intervention	Control	Intervention	Mortality	TTDA
0.75	1	36.6	27.5	20.8	20.8	93%	78%
0.75	1.1	36.6	27.4	21.8	19.8	94%	94%
0.75	1.2	36.6	27.4	22.7	18.9	93%	99%
0.75	1.3	36.6	27.4	23.4	18.1	93%	100%
0.8	1	35.5	28.5	20.8	20.8	75%	57%
0.8	1.1	35.7	28.3	21.8	19.8	78%	85%
0.8	1.2	35.5	28.4	22.7	18.9	75%	94%
0.8	1.3	35.5	28.4	23.5	18.1	75%	99%
0.85	1	34.5	29.5	20.8	20.8	46%	32%
0.85	1.1	34.6	29.4	21.8	19.8	50%	65%
0.85	**1.2**	**34.6**	**29.4**	**22.6**	**18.9**	**47%**	**83%**
0.85	1.3	34.7	29.5	23.5	18.1	49%	94%
0.9	1	33.7	30.3	20.8	20.8	25%	19%
0.9	1.1	33.7	30.3	21.8	19.8	23%	41%
0.9	1.2	33.7	30.3	22.8	18.9	24%	70%
0.9	1.3	33.7	30.3	23.5	18.1	24%	85%

Green good power, yellow acceptable power and red unacceptable power. Assumes two-sided tests at alpha=0.05. Assumes two-sided tests at alpha=0.05.

Table 3: Power with 750 Per Arm

Assumed Effect Size		60-Day Mortality		Median TTDA Among Survivors		Power	
Mortality HR	TTDA HR	Control	Intervention	Control	Intervention	Mortality	TTDA
0.75	1	36.6	27.4	20.8	20.8	97%	87%
0.75	1.1	36.6	27.4	21.8	19.8	96%	98%
0.75	1.2	35.5	27.5	22.6	18.9	96%	100%
0.75	1.3	36.6	27.4	23.6	18.1	96%	100%
0.8	1	35.5	28.5	20.8	20.8	84%	61%
0.8	1.1	36.6	28.4	21.8	19.8	84%	90%
0.8	1.2	36.6	28.4	22.7	18.9	85%	98%
0.8	1.3	36.6	28.4	23.5	18.1	85%	100%
0.85	1	34.6	29.3	20.8	20.8	60%	41%
0.85	1.1	34.6	29.4	21.8	19.8	58%	73%
0.85	1.2	34.6	29.4	22.7	18.9	58%	92%
0.85	1.3	34.6	29.5	23.5	18.1	58%	98%
0.9	1	33.7	30.3	20.8	20.8	28%	21%
0.9	1.1	33.7	30.4	21.8	19.8	30%	52%
0.9	1.2	33.7	30.3	22.8	18.9	28%	79%
0.9	1.3	33.7	30.4	23.5	18.1	27%	92%
1	1	32	32	20.8	20.8	5%	4%
1	1.1	32	32	21.8	19.8	6%	15%
1	1.2	32	32	21.8	18.9	5%	37%
1	1.3	32	32	23.6	18.1	6%	63%

Green good power, yellow acceptable power and red unacceptable power. Assumes two-sided tests at alpha=0.05. Assumes two-sided tests at alpha=0.05.

Table 4: Power with 1000 Per Arm

Assumed Effect Size		60-Day Mortality		Median TTDA Among Survivors		Power	
Mortality HR	TTDA HR	Control	Intervention	Control	Intervention	Mortality	TTDA
0.75	1	36.6	27.5	20.7	20.9	99%	93%
0.75	1.1	36.5	27.4	21.8	19.8	100%	100%
0.75	1.2	36.6	27.4	22.7	18.9	100%	100%
0.75	1.3	36.6	27.4	23.6	18.1	100%	100%
0.8	1	35.6	28.5	20.7	20.8	93%	77%
0.8	1.1	35.5	28.5	21.8	19.8	92%	97%
0.8	1.2	35.5	28.4	22.7	18.9	92%	100%
0.8	1.3	35.5	28.4	23.5	18.5	92%	100%
0.85	1	35.5	29.4	20.8	20.8	70%	49%
0.85	1.1	34.6	29.4	21.8	19.8	67%	83%
0.85	1.2	35.5	29.5	22.7	18.9	67%	96%
0.85	1.3	34.6	29.4	23.5	18.1	70%	100%
0.9	1	33.7	30.3	20.8	20.8	38%	25%
0.9	1.1	33.7	30.4	21.8	19.8	37%	63%
0.9	1.2	33.7	30.3	22.7	18.9	37%	87%
0.9	1.3	33.7	30.3	23.5	18.1	38%	98%
1	1	32	32	20.8	20.8	6%	6%
1	1.1	32	32	21.8	19.8	5%	16%
1	1.2	32	32	22.7	18.9	6%	45%
1	1.3	33.7	30.3	23.5	18.1	5%	76%

Green good power, yellow acceptable power and red unacceptable power. Assumes two-sided tests at alpha=0.05. Assumes two-sided tests at alpha=0.05.

9 Appendix B: Statistical Analysis Plan (SAP) Checklist v 1.0 2019

Section/Item	Index	Description	Reported on page #
Section 1: Administrative information			
Trial and Trial registration	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable)	1
	1b	Trial registration number	1
SAP Version	2	SAP version number with dates	1
Protocol Version	3	Reference to version of protocol being used	1
SAP revisions	4a	SAP revision history	1
	4b	Justification for each SAP revision	1
	4c	Timing of SAP revisions in relation to interim analyses, etc.	1
Roles and responsibility	5	Names, affiliations, and roles of SAP contributors	2
Signatures of:	6a	Person writing the SAP	1, 3
	6b	Senior statistician responsible	1
	6c	Chief investigator/clinical lead	1
Section 2: Introduction			
Background and rationale	7	Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial	6
Objectives	8	Description of specific objectives or hypotheses	6
Section 3: Study Methods			
Trial design	9	Brief description of trial design including type of trial (e.g., parallel group, multi-arm, crossover, factorial) and allocation ratio and may include brief description of interventions	6, 7

Randomization	10	Randomization details, e.g., whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)	7
Sample size	11	Full sample size calculation or reference to sample size calculation in protocol (instead of replication in SAP)	7
Framework	12	Superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis	8
Statistical interim analysis and stopping guidance	13a	Information on interim analyses specifying what interim analyses will be carried out and listing of time points	8
	13b	Any planned adjustment of the significance level due to interim analysis	8
	13c	Details of guidelines for stopping the trial early	8
Timing of final analysis	14	Timing of final analysis, e.g., all outcomes analysed collectively or timing stratified by planned length of follow-up	8
Timing of outcome assessments	15	Time points at which the outcomes are measured including visit “windows”	8
Section 4: Statistical Principals			
Confidence intervals and *P* values	16	Level of statistical significance	9
	17	Description and rationale for any adjustment for multiplicity and, if so, detailing how the type 1 error is to be controlled	8
	18	Confidence intervals to be reported	8
Adherence and Protocol deviations	19a	Definition of adherence to the intervention and how this is assessed including extent of exposure	8-9
	19b	Description of how adherence to the intervention will be presented	8-9
	19c	Definition of protocol deviations for the trial	8-9
	19d	Description of which protocol deviations will be summarized	8-9

Analysis populations	20	Definition of analysis populations, e.g., intention to treat, per protocol, complete case, safety	10
Section 5: Trial Population			
Screening data	21	Reporting of screening data (if collected) to describe representativeness of trial sample	10
Eligibility	22	Summary of eligibility criteria	10
Recruitment	23	Information to be included in the CONSORT flow diagram	10
Withdrawal/ Follow-up	24a	Level of withdrawal, e.g., from intervention and/or from follow-up	10
	24b	Timing of withdrawal/lost to follow-up data	10
	24c	Reasons and details of how withdrawal/lost to follow-up data will be presented	10
Baseline patient characteristics	25a	List of baseline characteristics to be summarized	10
	25b	Details of how baseline characteristics will be descriptively summarized	10
Section 6: Analysis			
Outcome definitions		List and describe each primary and secondary outcome including details of:	11
	26a	Specification of outcomes and timings. If applicable include the order of importance of primary or key secondary end points (e.g., order in which they will be tested)	11
	26b	Specific measurement and units (e.g., glucose control, hbA1c [mmol/mol or %])	NA
	26c	Any calculation or transformation used to derive the outcome (e.g., change from baseline, QoL score, Time to event, logarithm, etc.)	11-13
Analysis methods	27a	What analysis method will be used and how the treatment effects will be presented	11-13
	27b	Any adjustment for covariates	13
	27c	Methods used for assumptions to be checked for statistical methods	13-14
	27d	Details of alternative methods to be used if distributional assumptions do not hold, e.g., normality, proportional hazards, etc.	13-14

	27e	Any planned sensitivity analyses for each outcome where applicable	11-14
	27f	Any planned subgroup analyses for each outcome including how subgroups are defined	14
Missing data	28	Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)	14
Additional analyses	29	Details of any additional statistical analyses required, e.g., complier-average causal effect10 analysis	14
Harms	30	Sufficient detail on summarizing safety data, e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analysed, i.e., grade 3/4 only, incidence case analysis, intervention emergent analysis	NA
Statistical software	31	Details of statistical packages to be used to carry out analyses	15
References	32a	References to be provided for nonstandard statistical methods	16
	32b	Reference to Data Management Plan	NA
	32c	Reference to the Trial Master File and Statistical Master File	NA
	32d	Reference to other standard operating procedures or documents to be adhered to	NA

Taken from the paper: Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-43.

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; hbA1c, haemoglobin A1c; QoL, quality of life; SAP, statistical analysis plan.

1.2 Signatures

I have read and approve the enclosed SAP dated 2022-03-9 for EFFORT trial.

Senior Statistician & SAP Author

Name: Andrew G. Day

Signature: Andrew G. Day

Date: March 9, 2022

Statistician Performing Analysis (other than senior statistician):

Name: Xuran Jiang

Signature: Xuran Jiang

Date: 10Mar2022

Trial Co-ordinator

Name: Jennifer Korol

Signature: JKorol

Date: March 10, 2022

Principal Investigator

Name: Daren K. Heyland

Signature:

Date: Mar 9/2022