ClinicalTrials.gov Identifier: NCT04350593

Statistical Analysis Plan	
Study Code	D1690C00081
Edition Number	5.0
Date	3 March 2021

An International, Multicenter, Randomized, Double-blind, Placebo-controlled, Phase III Study Evaluating the Efficacy and Safety of Dapagliflozin in Respiratory Failure in Patients with COVID-19

Short Title: DARE-19 (Dapagliflozin in Respiratory failure in patients with COVID-19)

Based on Clinical Study Protocol, version 4.0, 20 November 2020

An International, Multicenter, Randomized, Double-blind, Placebo-controlled, Phase III Study Evaluating the Efficacy and Safety of Dapagliflozin in Respiratory Failure in Patients with COVID-19

Short Title: DARE-19 (Dapagliflozin in Respiratory failure in patients with COVID-19)

An International, Multicenter, Randomized, Double-blind, Placebo-controlled, Phase III Study Evaluating the Efficacy and Safety of Dapagliflozin in Respiratory Failure in Patients with COVID-19

Short Title: DARE-19 (Dapagliflozin in Respiratory failure in patients with COVID-19)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
ALT	Alanine aminotransferase
ATC	Anatomical Therapeutic Chemical
BiPAP	Bilevel positive airway pressure
CKD-EPI	Chronic kidney disease epidemiology collaboration equation
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
CPAP	Continuous positive airway pressure
CSP	Clinical study protocol
DAE	Adverse event leading to treatment discontinuation
DKA	Diabetic ketoacidosis
ECMO	Extracorporeal membrane oxygenation
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
FAS	Full analysis set
HF	Heart failure
HR	Hazard ratio
hsCRP	High sensitivity C-reactive protein
ICF	Informed consent form
ICU	Intensive care unit
IP	Investigational Product (dapagliflozin or matching placebo)
ITT	Intention-to-treat
KM	Kaplan-Meier
LDH	Lactate dehydrogenase
LTFU	Lost to follow-up
MedDRA	Medical Dictionary for Regulatory Activities
NEWS 2	National Early Warning Score 2
NT-proBNP	N-terminal pro b-type natriuretic peptide
PCR	Polymerase chain reaction
PT	Preferred term

Abbreviation or special term	Explanation
SAE	Serious adverse event
SAS	Safety analysis set
SGLT2i	Sodium-glucose co-transporter 2 inhibitors
SOC	System organ class
T2D	Type 2 diabetes
WoC	Withdrawal of consent
WR	Win Ratio

AMENDMENT HISTORY

Date /Version	Brief description of change
14 April 2020 / Version 1.0	Version 1.0 signed

1 August 2020 / Version 2.0	The number of sites has been updated to 90, the list of countries has been changed as well [section 1.2.1] Extended follow-up of 60 days is added in accordance with the CSP updated version 3.0 [section 1.2.1] "before Day 30" is replaced by "after Day 30" [section 1.2.4] Recruitment assumption is removed, since it does not have an impact on power calculations. Previous version contained estimate of 800 patients needed with 900 being randomized. In the current version the estimate of 800 patients has been removed and only the number 900 is retained. [section 1.4] Section on Estimands has been added [section 3.1] Censoring rule for the primary endpoint has been added [section 3.2] Censoring rule for the discharge from hospital has been modified to censor patients who die in hospital not at the time of death but at Day 30 [section 3.2] The endpoint "Total Number of Days Alive, Out of Hospital and/or Free from Mechanical Ventilation" is clarified as "Total Number of Days Alive and Free from Mechanical Ventilation" [section 3.3.2] The definition of acute kidney injury has been updated to included events happening in outpatient setting [section 3.3.1] The endpoint of acute kidney injury has been updated to include death as a composite measure to account for intercurrent event due to death [section 3.3.1] Updated to state that new/worsened organ dysfunction will not be included in the hierarchical testing, but analyzed as a component of the primary endpoint. As such repeated details about its derivation were moved to section 3.2 Events of acute renal failure have been added to safety analysis [section 3.4] Clarification of calculation of study drug compliance added. Calculation of proportion of patient-days on study drug has been added [section 4.1] The fixed-sequence of testing the secondary endpoints has been clarified. New/worsened organ dysfunction is removed from the testing hierarchy [section 4.1.2] Detailed definition of prior, baseline and concomitant medications has been added [section 4.2.2] Use of proportionality assumption has been clarified and the handling of ties has been specified in the Cox regression analysis [section 4.2.3] A sensitivity analysis of the primary endpoint is redefined to exclude patients who were not tested for SARS-CoV-2 at randomization and tested negative when testing became available [section 4.2.4.3] Details for the tipping point analysis has been added [section 4.2.4.3] Detailed description of analysis methods for the Total Number of Days endpoints has been added [section 4.2.6.2] Clarification of definitions of time-to-event variables in testing hierarchy has been added [section 4.2.6] Actual exposure is introduced [section 4.2.7]

Date /Version	Brief description of change
	Section on laboratory evaluations has been added [section 4.2.7.4] Marked abnormalities have been defined [section 4.2.7.5]
1 November 2020 / Version 3.0	Definition of the composite strategy of handling intercurrent events has been added [section 3.1] A sensitivity analysis has been added to incorporate into primary endpoint deaths happening after incomplete in-hospital event assessment [section 3.2] Safety Data Restriction Specification has been added to restrict analysis of safety objective only to 30-day treatment period [section 3.4.1] Clarification for derivation of patient-days with complete follow-up has been added [section 4.1.4] Sensitivity analysis for the difference in proportions has been added for the primary endpoint [section 4.2.4.3] Details of the tipping point analysis have been added [section 4.2.4.3] Sensitivity analysis for the difference in proportions has been added for the time-to-event endpoints [section 4.2.6] Figure 2 has been added for the interpretation of the hierarchical composite endpoint (HCE) [section 4.2.5] Derivation of the analysis values for the HCE analysis has been added [section 4.2.5] Handling of missing data for the HCE analysis has been added [section 4.2.5] Sensitivity analysis for “Total Number of Days” endpoints has been added [section 4.2.6.2] Analysis of the observational period has been specified. [section 6] Details of the win ratio confidence interval calculation have been added [APPENDIX]

Date /Version	Brief description of change
1 February 2021/ Version 4.0	Primary, Secondary, Safety and Exploratory objectives are updated based on the protocol amendment [sections 1.1.1, 1.1.2, 1.1.3, 1.1.4]
	Event driven study has been changed to a fixed sample size study, and sample size increased to approximately 1200 patients, based on the protocol amendment [section 1.2.1]
	Section is added to specify the data collection during the extended follow-up period [section 1.2.6]
	Sample size calculation has been changed to reflect the updated protocol accounting for the dual primary endpoints [section 1.4]
	Estimands, Primary variables, Secondary variables are updated to reflect the dual endpoint nature of the study, change of the Acute Kidney Endpoint to Composite Kidney Endpoint, according to the protocol amendment, and to reflect the changed order in testing sequence of secondary endpoints [section 3.1, 3.2, 3.3.1-3.3.5]
	Pooling of countries for stratified analysis has been specified [section 4.1]
	The definition of compliance has been amended to distinguish between compliant and compliant while alive [section 4.1]
	Confirmatory testing procedure has been updated [section 4.1.2]
	Two additional supplementary analyses are added for the prevention composite endpoint – multiple event analysis and removing acute kidney injury from the composite [section 4.2.4.2]
	Sensitivity analysis is added for the prevention composite endpoint by baseline use of remdesivir [section 4.2.4.3]
	The hierarchical composite endpoint section has been amended to include the stratified log-rank test as the primary method of analysis and the win ratio from the stratified Cox regression as the main method of treatment effect estimate applied to the HCE [section 4.2.5]
	The hierarchical composite endpoint section has been amended to include the definition of recovery, sensitivity analyses by excluding patients with unconfirmed SARS-CoV-2 test or by baseline use of remdesivir [section 4.2.5]
	Section “Changes from protocol” has been removed, since after the protocol amendment there are no changes from the protocol.
	Software implementation of the multi-state fallback procedure has been provided [APPENDIX]

Date /Version	Brief description of change
3 March 2021/ Version 5.0	In the confirmatory testing procedure "Death from any cause" moved up the hierarchy [section 4.1.2] Clarification has been added about reporting the individual events contributing to the composite endpoint if several of these components occur on the same day [section 4.2.4.2] The sensitivity analysis which includes deaths after incomplete event assessment has been added to section 4.2.4.3 in addition to being mentioned in section 3.2.1 It has been clarified that type a) missingness will be handled for the primary analysis of the hierarchical composite endpoint [section 4.2.5.3] The primary analysis method of the "Total number of days" endpoints has been updated to incorporate hierarchical structure of the outcomes (death vs survival). Analysis also incorporates patients with premature study discontinuation as censored [section 4.2.6.2] Software implementation of the primary analysis for HCE has been provided [APPENDIX]

1 STUDY DETAILS

1.1 Study Objectives

1.1.1 Primary Objective

Primary objective:	**Outcome measure:**
To determine whether dapagliflozin 10 mg is superior to placebo, in reducing disease progression, complications, and all-cause mortality in patients hospitalized with COVID-19.	Dual primary endpoints of: **Prevention of COVID-19 complications or death** During the 30-day treatment period, time to first occurrence of new/worsened organ dysfunction during index hospitalization or death from any cause. New/worsened organ dysfunction is defined as at least one of the following: • Respiratory decompensation requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP), and/or initiation of ECMO • New or worsening congestive HF[a] • Requirement for vasopressor therapy and/or inotropic or mechanical circulatory support • Ventricular tachycardia or fibrillation lasting at least 30 seconds and/or associated with hemodynamic instability or pulseless electrical activity, or resuscitated cardiac arrest • Doubling of s-Creatinine or initiation of renal replacement therapy **Improving clinical recovery** Hierarchical composite outcome measure: 1 Death from any cause through Day 30 2 New/worsened organ dysfunction (as defined above) 3 Clinical status at Day 30 for patients still hospitalized and without any worsening organ dysfunction (using points 3 to 5 of a 7-point ordinal scale[b]) 4 Hospital discharge before Day 30 and alive at Day 30

[a] Congestive HF is defined as at least one of the following 1) initiation of new intravenous therapy for heart failure 2) reinstitution of previous intravenous therapy for heart failure 3) increase in current intravenous therapy for heart failure. This is based on modification on previous definition of in-hospital worsening heart failure (McMurray et al 2007)

[b] 7-point Patient Clinical Status scale:

1 Not hospitalized, no limitations on activities
2 Not hospitalized, limitation on activities

3 Hospitalized, not requiring supplemental oxygen
4 Hospitalized, requiring supplemental oxygen
5 Hospitalized, on high flow oxygen devices
6 Hospitalized, on invasive mechanical ventilation or ECMO
7 Death

BiPAP Bilevel positive airway pressure; COVID-19 Coronavirus disease 2019; CPAP Continuous positive airway pressure; ECMO extracorporeal membrane oxygenation; HF Heart failure.

1.1.2 Secondary Objectives

Secondary objectives:	Outcome measures:
To compare the effect of dapagliflozin 10 mg versus placebo on time to hospital discharge.	Time to hospital discharge[a]
To compare the effect of dapagliflozin 10 mg versus placebo on total number of days alive and free from respiratory decompensation requiring mechanical ventilation	Total number of days alive and free from respiratory decompensation[a] requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP) from randomization through Day 30
To compare the effect of dapagliflozin 10 mg versus placebo on total number of days alive, not in ICU, and free from respiratory decompensation requiring mechanical ventilation	Total number of days alive, not in the ICU, and free from respiratory decompensation[a] requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP) from randomization through Day 30
To compare the effect of dapagliflozin 10 mg versus placebo on a composite kidney endpoint	Time to composite of acute kidney injury[b] or initiation of renal replacement therapy[c], or death from any cause through Day 30
To compare the effect of dapagliflozin 10 mg versus placebo in reducing the incidence of all-cause mortality	Time to death from any cause through Day 30

[a] Refers to index hospitalization only

[b] Acute kidney injury defined as:

- An episode of doubling s-Creatinine compared to baseline during index hospitalization
- or SAE with preferred term of Acute kidney injury following discharge and through Day 30

[c] Renal replacement therapy defined as:

- Initiation of renal replacement therapy during index hospitalization
- or SAE with a preferred term for renal replacement therapy (ie, Haemodialysis, Haemofiltration, Continuous haemodiafiltration, Dialysis, Peritoneal dialysis, Dialysis device insertion, Renal replacement therapy, or Artificial kidney device user) following discharge and through Day 30

BiPAP Bilevel positive airway pressure; COVID-19 Coronavirus disease 2019; CPAP Continuous positive airway pressure; ECMO Extracorporeal membrane oxygenation; ICU Intensive care unit; SAE Serious adverse event.

1.1.3 Safety Objectives

Safety objective:	**Outcome measures:**
To evaluate the safety and tolerability of dapagliflozin compared to placebo in patients hospitalized with COVID-19.	• Serious adverse events from randomization to Day 30[a] • Acute kidney injury defined as: ° An episode of doubling of s-Creatinine compared to baseline during index hospitalization ° or SAE with preferred term of acute kidney injury following discharge and through Day 30 • Incidence of diabetic ketoacidosis from randomization through Day 30

[a] SAEs will be collected through Day 90 but comparison of treatment groups will be assessed based on the data obtained through Day 30. For the definition of reportable SAEs, see CSP Section 6.4.2.

COVID-19 Coronavirus 2019; SAE Serious adverse event.

1.1.4 Exploratory Objectives

Exploratory objective:	**Outcome measures:**
To compare the effect of dapagliflozin 10 mg to placebo on the components of the primary endpoint, biomarkers, and patient's clinical status.	• Change in NT-proBNP, hs troponin, D-dimer, LDH, ALT, lymphocyte count, CRP between Day 1 and Day 15 (or discharge from hospital, whichever is earlier) • Qualitative PCR for SARS-CoV-2 in oropharyngeal/nasopharyngeal swab at baseline (while hospitalized); and Day 15 (if still hospitalized) or discharge from hospital • Change in NEWS 2 from Day 1 to Day 15 (or discharge from hospital, whichever is earlier). • Patient's clinical status (on a 7-point ordinal scale) at Day 15 (or discharge from hospital, whichever is earlier) • Total number of days alive and not on renal replacement therapy[a] • Proportion of patients with acute coronary syndrome[b]

[a] Refers to index hospitalization only

[b] Acute coronary syndrome defined as: abnormal troponin level above 99th percentile of the local laboratory reference range or, if abnormal at baseline, further rise in troponin levels accompanied by at least 1 of the following: 1) ischemic symptoms 2) ischemic ST-segment changes on ECG (Thygesen et al 2018)

NEWS 2 is a standardized assessment of acute-illness severity and can prompt critical care intervention. It is used as an adjunct to clinical judgment.

ALT Alanine aminotransferase; COVID-19 Coronavirus disease 2019; CRP C-reactive protein; ECG Electrocardiogram; hs troponin High-sensitivity cardiac troponin; LDH lactate dehydrogenase; NEWS 2 National Early Warning Score 2; NT-proBNP N-terminal-pro B-type natriuretic peptide; PCR Polymerase chain reaction SARS-CoV-2 Severe acute respiratory syndrome-coronavirus-2.

1.2 Definitions

1.2.1 Study Closure

The study is a fixed follow-up study with follow-up time of 30 days. Approximately 1200 patients will be randomized at approximately 90 sites in North America (USA and Canada), Latin America (Argentina, Brazil, Mexico), India, and Europe (UK).

An extended follow-up period of an additional 60 days of observational follow up (on top of the current active treatment duration of 30 days) will be conducted to examine any potential longer-term trajectory of recovery from COVID-19 among trial participants. All analyses described in the current SAP (except Section 6) will be carried out on data collected up to and including Day 30. As soon as the pre-planned number of patients have completed their 30 day treatment period, and the data is collected and cleaned, the database will be locked and unblinding performed for these analyses.

Separate analyses will be done on extended follow-up period described in Section 6.

1.2.2 Withdrawal of Informed Consent

Withdrawal of consent (WoC) should only occur if the patient has received appropriate information about, and does not agree to, any kind of further assessments or contact, including modified follow-up options. A patient who considers withdrawing from the study must be informed by the Investigator about modified follow-up options (eg, telephone contact, a contact with a relative or treating physician, or information from medical records).

If the patient withdraws consent for disclosure of future information, the Sponsor may retain and continue to use any data collected before such a WoC. A patient who withdraws consent will always be asked about the reason(s) and the presence of any adverse event (AE). The Investigator will follow up patients as medically indicated. To ensure validity of study data, efforts will be made to collect as much data as possible throughout the study (including after hospital discharge) and especially vital status (dead or alive) (also for patients who have withdrawn their informed consent). Therefore, if informed consent has been withdrawn completely or the patient is non-contactable following hospital discharge, the Investigator will attempt to collect information on all patients' vital status from publicly available sources at study closure, in compliance with local privacy laws/practices.

1.2.3 Discontinuation from Study Drug

Discontinuation from investigational product (IP) is not the same as complete withdrawal from the study. A patient who discontinues IP should optimally continue to follow up study assessments up to and including Day 90. Alternatively, if the patient does not agree to this approach, modified follow-up should be arranged (eg, less frequent assessments, one contact at Day 30, or other means). Patients who agree to some kind of modified follow-up are still participating in the study. The modified visits and procedures that are done will be recorded in the electronic case report form (eCRF).

Data from patients who did not withdraw consent or events occurring before WoC for patients who withdraw consent will be included in the intention-to-treat (ITT) analyses irrespective of whether the event occurred before or following discontinuation of study drug.

1.2.4 Vital Status

Known vital status at the end of the 30-day treatment period will be defined when the patient is dead or has date last known alive on or after Day 30. Vital status for 60 and 90 days will be summarized separately.

For patients who have withdrawn consent, the investigator will attempt to collect vital status from publicly available sources at study closure in compliance with local privacy laws/practices.

1.2.5 Follow-up Period after Hospital Discharge

Patients discharged from hospital will continue with daily treatment for the remainder of the 30 days, and are followed up by telephone at Day 15 and/or Day 30 (± 3 days).

The following assessments will be completed:

- Concomitant medications will be recorded
- Follow up of ongoing reportable serious AEs (SAEs). Details of any reportable SAEs will be recorded.
- Vital status
- Investigational product adherence
- Details of any re-hospitalization for the patient
- Patient Clinical Status will be assessed using a 7-point scale (defined in Section 1.1.2)

1.2.6 Extended Follow-up Period after Day 30

Following last dose of investigational product at Day 30, patients are followed up by telephone at Day 60 and Day 90 (± 7 days).

The following assessments will be completed:

- Concomitant medications will be recorded

- Follow up of ongoing reportable SAEs. Details of any new reportable SAEs will be recorded.
- Vital status
- Details of any re-hospitalization for the patient
- Patient Clinical Status will be assessed using a 7-point scale (defined in Section 1.1.2)

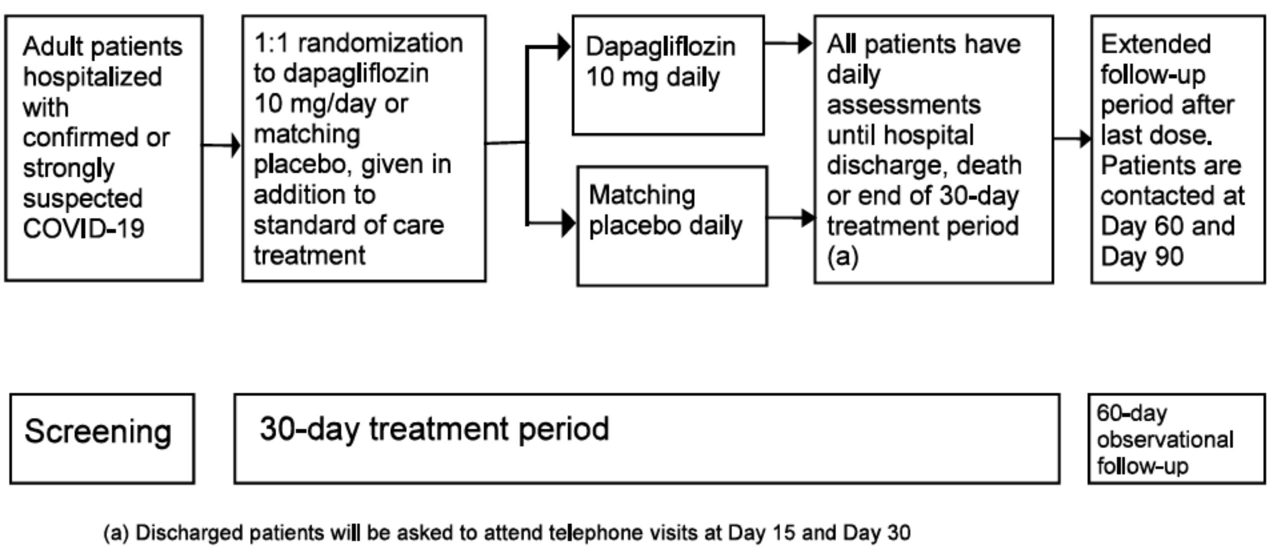
1.2.7 Lost to Follow-up

The term lost to follow-up (LTFU) will be limited to patients with unknown vital status at the end of the study as defined in Section 1.2.4. Other measures will be used to describe incomplete follow-up of the prevention endpoint (Section 4.1.4).

1.3 Study Design

This is an international, multicenter, parallel-group, randomized, double-blind, placebo-controlled, phase III study in approximately 1200 hospitalized adult COVID-19 patients. The study is evaluating the effect of dapagliflozin 10 mg versus placebo, given once daily for 30 days in addition to background local standard of care therapy, including treatments to control co-morbidities. The screening period should be as short as possible (no more than 2 days), and patients should start investigational product the same day as randomization. Following last dose of investigational product at Day 30 patients are followed up for an extended additional (observational) period of 60 days.

Figure 1 - Study flow chart



1.3.1 Randomization

In this study, patients will be recruited from sites with adult patients hospitalized with confirmed or strongly suspected COVID-19. Each patient should meet all of the inclusion criteria and none of the exclusion criteria for this study in order to be randomized to investigational product. Under no circumstances can there be exceptions to this rule. Patients can be re-screened once, but must be considered screen failures if they still do not meet the entry requirements (clinical study protocol [CSP], Section 3.3).

In this study, "enrolled" patients are those who sign the informed consent form (ICF). "Randomized" patients are those who undergo randomization and receive a randomization number.

All patients will be randomly assigned to investigational product centrally using an Interactive Response Technology system. Randomization to investigational product will be performed in balanced blocks to ensure approximate balance between the treatment groups (1:1). The Sponsor or delegate is responsible for generating the randomization scheme for this study using a validated system. Before the study starts, the instructions for accessing and using the Interactive Response Technology system will be provided to each site.

Investigational product (dapagliflozin or placebo) should be administered the same day the IP kit number is assigned and as soon after randomization as possible.

To ensure balanced randomization within countries, stratification will be employed by country.

1.4 Number of Patients

The primary objectives of the study are to determine the superiority of dapagliflozin versus placebo in reducing the incidence of complications or all-cause mortality (prevention of worsening COVID-19) or improving clinical recovery. It is estimated that a sample size of approximately 1200 patients will provide adequate power to detect the treatment effect on prevention or recovery, when the dual primary endpoints are used for testing, with alpha split between these endpoints.

Since the original protocol was designed, the unpredictable nature of the evolving global pandemic and the change in standard of care for treatment of COVID-19 resulted in lower than expected event rates. As a consequence, faster and more complete recovery has now become an important treatment goal on par with prevention of complications and death in patients hospitalized with COVID-19, prompting the addition of 'recovery' to the primary objectives.

The initial version of the protocol (CSP version 1.0, 2 April 2020) specified an event-driven approach with 380 events needed to detect HR of 0.75 with 80% power. In a fixed follow-up study that would have required 42% of initially randomized 900 patients to experience an event. The provision for a potential increase of sample size was included in the initial design and intended as a way of balancing the possible decrease in event rates, while the recruitment rate of 900 patients was anticipated to occur over a period of approximately 3 months. When the CSP was updated (version 4.0, 20 November 2020) it was estimated that around 10 to 20% of patients will develop COVID-19 related complications during the index hospitalization or will experience death during the 30-day treatment period, while 80 to 90% of patients will recover without experiencing worsening. Therefore, the sample size of approximately 1200 patients would provide approximately 100 to 250 events for the first dual primary endpoint (prevention).

With dual primary objectives of prevention of COVID-19 complications or death and improvement in clinical recovery, an effect on any one will be sufficient evidence of the effectiveness of study medication. This is per the FDA guidance on Multiple Endpoints in Clinical Trials (FDA 2017). To control the type I error for dual primary endpoints, the allocated alpha of 5% will be split between them. Table 1 below shows the true hazard ratio required for 80% power for a hypothetical scenario of an even split of alpha (2.5% two-sided for each primary endpoint) depending on the number of events observed (final alpha allocation is described in Section 4.1.2). Since no prior studies are available for SGLT2 inhibitors in the COVID-19 setting, possible scenarios of event rates and treatment effects are considered to infer the sample size (Table 1).

Table 1 - 2.5% Alpha (80% Power, Time-to-event Analysis)

Number of Events	Hazard Ratio (Dapagliflozin versus Placebo) required for 80% power	Minimal Detectable Hazard Ratio
100	0.54	0.64
150	0.6	0.69
200	0.65	0.73
250	0.68	0.75

For the second primary endpoint (recovery) the sample size of 1200 patients will detect a win ratio (WR) of 1.23 with at least 80% power for hypothetical alpha of 2.5% (The power is calculated based on asymptotic normality property of the win proportion [WP] where WR = WP/[1-WP] and the estimated standard deviation (SD) for the WP is assumed to be

SD = 1/sqrt[3] = 0.57735 [Kawaguchi et al 2011], as it is a conservative estimate.) This is based on an overall 1:1 allocation between dapagliflozin and placebo.

2 ANALYSIS SETS

2.1 Definition of Analysis Sets

2.1.1 Full Analysis Set

All patients who have been randomized to study treatment will be included in the full analysis set (FAS) irrespective of their protocol adherence and continued participation in the study. Patients will be analyzed according to their randomized IP assignment, irrespective of the treatment actually received. The FAS will be considered the primary analysis set for the primary, secondary, and exploratory efficacy variables.

2.1.2 Safety Analysis Set

All randomized patients who received at least 1 dose of randomized treatment will be included in the safety analysis set (SAS). Patients will be analyzed according to the treatment actually received. For any patients given incorrect treatment, ie, randomized to one of the treatment groups but actually given the other treatment, the treatment group will be allocated as follows: patients who got both incorrect and correct treatment will be analyzed according to their randomized treatment; patients who got only the incorrect treatment will be analyzed according to that treatment.

The SAS will be considered the primary analysis set for all safety variables.

2.2 Violations and Deviations

The important protocol deviations listed below will be summarized by randomized treatment group

- Patients who were randomized but did not meet inclusion and exclusion criteria.
- Patients who received the wrong study treatment at any time during the study.
- Patients who received prohibited concomitant medication, which for this study is limited to open-label SGLT2 inhibitors taken in combination with IP.

As the primary analysis is an ITT analysis, a protocol deviation will not imply exclusion from the primary analysis.

3 PRIMARY AND SECONDARY VARIABLES

3.1 Estimands

The treatment in this clinical trial is dapagliflozin added to background local standard of care therapy, compared to standard of care alone. The population will include all patients on low-flow oxygen with COVID-19 and risk factors, as represented by the inclusion/exclusion criteria. The primary treatment effect is characterized as risk reduction of complications and all-cause mortality (prevention), or improvement in clinical recovery. Specifically, the primary variable of interest is the occurrence of COVID-19 related new/worsened organ dysfunction from randomization through hospital discharge (ie, during index hospitalization) or death from any cause through Day 30, which will have the same definition in both primary endpoints.

Intercurrent events are events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest (ICH E9(R1) 2019). Intercurrent events such as treatment discontinuation, or changes or use of additional medication may occur in this trial. Our plan is to apply the treatment policy strategy by using all values of the variables of interest regardless of the occurrence of these intercurrent events in the analysis.

In addition, for the prevention composite endpoint, hospital discharge is an intercurrent event, since it affects the measurements of the components of the composite endpoint as only vital status is used after hospital discharge from the components of the prevention endpoint. This intercurrent event will be handled using the treatment policy strategy, that is the occurrence of the intercurrent event is considered irrelevant in defining the treatment effect.

For the hierarchical composite endpoint (recovery), the same intercurrent event will be handled using the composite strategy by including it in the definition of the endpoint, because the HCE is meant to capture the treatment effect on recovery, hence the occurrence of hospital discharge is informative. The guidance ICH E9(R1) 2019 defines the composite strategy as

"An intercurrent event is considered in itself to be informative about the patient's outcome and is therefore incorporated into the definition of the variable."

For the kidney endpoint, we apply the composite estimand strategy and include death in the composite.

3.2 Primary Variables

3.2.1 Prevention Composite Endpoint

The analysis of the prevention endpoint is based on time from randomization to first occurrence of new/worsened organ dysfunction during index hospitalization or death from any cause. New/worsened organ dysfunction is defined as at least one of the following:

- Respiratory decompensation requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP), and/or initiation of ECMO
- New or worsening congestive HF
- Requirement for vasopressor therapy and/or inotropic or mechanical circulatory support
- Ventricular tachycardia or fibrillation lasting at least 30 seconds and/or associated with hemodynamic instability or pulseless electrical activity, or resuscitated cardiac arrest
- Doubling of s-Creatinine or initiation of renal replacement therapy

Congestive HF is defined as at least one of the following 1) initiation of new intravenous therapy for HF 2) reinstitution of previous intravenous therapy for HF 3) increase in current intravenous therapy for HF. This is based on modification on previous definition of in-hospital worsening HF (McMurray et al 2007).

The primary variable of interest is the onset of COVID-19 related new or worsened organ dysfunction in the hospital setting (ie, during index hospitalization), or death from any cause at any time during the 30-day treatment period. Therefore, if a patient is discharged from the hospital without an event of organ dysfunction, then the patient is assumed to have made a partial or full recovery from COVID-19 and is therefore no longer at high risk of developing COVID-19 related organ dysfunction in the outpatient setting. Thus, for the primary analysis (for both dual endpoints), after discharge only vital status will be considered in the analysis.

For the prevention endpoint the censoring rule is the following, if a patient is discharged from the hospital without an event of organ dysfunction and is alive at Day 30, then this patient is censored at Day 30. The only instances that the censoring date will be truncated is when patients have incomplete vital status assessment after hospital discharge (are lost to follow-up): the censoring date will be the date of hospital discharge or the last assessment of vital status after the discharge or the date of withdrawal of consent, if after withdrawal of consent the vital status is unknown.

Incomplete event assessment during the index hospitalization will serve as an additional censoring event. If it is known that some patients have died after incomplete event assessment, then a sensitivity analysis will be conducted to consider these deaths as events.

The number and percentage of patients with complete follow-up of the prevention endpoint will be reported.

In the analysis of components of the composite endpoint, date of death from any cause will be an additional point of censoring.

3.2.2 Hierarchical Composite Endpoint

The efficacy variable is a composite of ordinal outcomes. All patients will be ranked based on the timing and severity of their events. Up to Day 30, all events described in the definition of the composite will be considered (not just the first events) to determine the ranks of patients as described in Section 4.2.5.

3.3 Secondary Variables

The secondary endpoints are included in a hierarchical testing sequence following the primary endpoint as described in 3.3.1 to 3.3.5.

3.3.1 Composite Kidney Endpoint

The efficacy analysis is based on time from randomization to composite of acute kidney injury, initiation of renal replacement therapy or death from any cause.

Acute kidney injury is defined as:

- An episode of doubling s-Creatinine compared to baseline during index hospitalization
- or SAE with preferred term of Acute kidney injury following discharge and through Day 30

Renal replacement therapy is defined as:

- Initiation of renal replacement therapy during index hospitalization
- or SAE with a preferred term for renal replacement therapy (ie, Haemodialysis, Haemofiltration, Continuous haemodiafiltration, Dialysis, Peritoneal dialysis, Dialysis device insertion, Renal replacement therapy, or Artificial kidney device user) following discharge and through Day 30

3.3.2 Total Number of Days Alive and Free from Mechanical Ventilation

The efficacy variable is the total number of days alive and free from respiratory decompensation (during index hospitalization only) requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP) from randomization through Day 30.

The total number of days will be calculated as follows. The total follow-up time is defined for each patient as the 30 days after the randomization or until time of WoC or last contact for patients who are LTFU. The total time spent in hospital will be divided into combined duration of use of mechanical ventilation and of being free from mechanical ventilation. The time interval of being out of hospital and alive is considered as being free from mechanical ventilation. If a patient dies, the number of days from their death to the end of study will be assigned as days dead. Days in hospital with use of mechanical ventilation and days dead will then be subtracted from total follow-up time to arrive at days alive and free from mechanical ventilation for each patient.

3.3.3 Total Number of Days Alive, Not in the ICU, and Free from Mechanical Ventilation

The efficacy variable is the total number of days alive, not in the ICU, and free from respiratory decompensation (during index hospitalization only) requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP) from randomization through Day 30.

The total number of days will be calculated as follows. The total follow-up time is defined for each patient as the 30 days after the randomization or until time of WoC or last contact for patients who are LTFU. The total time spent in hospital will be divided into combined duration of being in ICU or using mechanical ventilation and of being out of ICU and free from mechanical ventilation. The time interval of being out of hospital and alive is considered as being out of ICU and free from mechanical ventilation. If a patient dies, the number of days from their death to the end of study will be assigned as days dead. Days in ICU or using mechanical ventilation and days dead will then be subtracted from total follow-up time to arrive at days alive, not in the ICU, and free from mechanical ventilation.

3.3.4 Death from Any Cause

The efficacy variable is time to from randomization to death from any cause. All deaths on or prior to Day 30, including death after WoC will be included. Patients who are alive will be censored at the earliest of date last known alive and Day 30.

3.3.5 Discharge from Hospital

The efficacy variable is time from randomization to discharge from hospital. Since death is a competing event for the hospital discharge, censoring of patients at the time of death is not appropriate. Instead patients dying before hospital discharge will be censored at the end of follow-up, that is at day 30 (Dodd et al. 2011). Hospital discharge is a desirable outcome and intention of the treatment is to reduce (as opposed to time to negative outcomes, for example, organ worsening) the time to this event. Therefore, it is reasonable to assume that patients dying in hospital did not have the event during the entire follow-up period, since death precludes the happening of hospital discharge.

3.4 Safety Variables

The safety and tolerability of dapagliflozin will be evaluated from reportable SAEs (see CSP Section 6.4.2), and safety events (acute kidney injury, DKA). Additionally, DAEs and events of "Acute renal failure", defined by narrow SMQ (Standardized MedDRA Queries) scope, will be summarized. Routine AEs will not be collected.

3.4.1 Safety Data Restriction Specification

Because of the observational additional 60 days of follow-up, some AEs emerging during the 30 days of follow-up will have resolution after Day 30. For analysis of all safety objectives only data from 30-days treated period will be used. Therefore, AEs still ongoing at the end of 30 Day treatment period with outcome after day 30 will be considered as ongoing. For the analysis of observational data see Section 6.

4 ANALYSIS METHODS

4.1 General Principles

No multiplicity adjustment will be made to confidence intervals as they will be interpreted descriptively and used as a measure of precision. P-values for variables not included in the confirmatory testing sequence or following a non-significant test in the sequence will be regarded as nominal.

For stratified analyses countries contributing less than 5% of all randomized patients will be pooled together.

Baseline Laboratory Value

For all laboratory variables, the baseline value is defined as the last value on or prior to date of randomization. Preference will be given to study-mandated laboratory values over standard of care laboratory values.

eGFR

The estimated glomerular filtration rate (eGFR) values will be calculated (in mL/min/1.73 m^2) from the creatinine measurements using the chronic kidney disease epidemiology collaboration equation (CKD-EPI) formula (Levey at al 2009). Descriptive statistics will be presented based on laboratory data.

Study Drug Compliance

The percentage of study drug compliance (compliant and compliant while alive) for the overall treatment period will be derived. Compliant is defined as continuing taking the investigational product until and on Day 30, or died before Day 30 while being on study treatment. Compliant while alive is defined as being compliant and alive at Day 30. Additionally, patient-days on drug will be defined as days from first dose date until the earliest of drug discontinuation, death, WoC, or Day 30. Proportion of patient-days on study drug will be calculated as total patient-days on study drug divided by total patient-days with maximum days on study drug, that is, time to death or Day 30.

4.1.1 Hypotheses

For the primary endpoints the following 2 hypotheses will be tested with alpha allocated to each hypothesis to maintain an overall 5% 2-sided significance level:

Prevention

H0:HR [dapagliflozin:placebo] = 1

versus

H1:HR [dapagliflozin:placebo] ≠ 1

Recovery

H0:WR [dapagliflozin:placebo] = 1

versus

H1:WR [dapagliflozin:placebo] ≠ 1

Here HR (hazard ratio) is estimated from a Cox regression stratified by (pooled) countries and adjusted for sex and age, while WR (win ratio) will be estimated from the Cox regression applied to ranks of the Hierarchical Composite Endpoint (HCE, Section 4.2.5). Note that the statistical test for the latter will be done using the stratified log-rank test. A direct win ratio estimate will be provided as well, based on pairwise comparisons of patients in the active group with the patients in the placebo group, as a supplementary analysis.

A strong type 1 error control will be applied in testing the primary and secondary efficacy endpoints (Section 4.1.2).

4.1.2 Confirmatory Testing Procedure

4.1.2.1 Multi-stage fallback procedure

To control the overall type I error at a 2-sided $\alpha=0.05$ level for multiplicity across primary and secondary endpoints (Table 2), a multi-stage fallback testing procedure (Dmitrienko et al 2006) will be employed. Graphical representation of this testing procedure is given in Figure 2 using the Bonferroni-based recycling framework developed in Burman et al 2009. Different weights will be allocated to the significance level for each hypothesis test in Table 2.

Table 2 - Multiple testing procedure

Hypothesis	Endpoint
1	The prevention composite endpoint (first occurrence of new/worsened organ dysfunction during index hospitalization or death from any cause through day 30).
2	The hierarchical composite endpoint (recovery).
3	The composite kidney endpoint.
4	Death from any cause.
5	Total number of days alive and free from respiratory decompensation.
6	Total number of days alive, not in the ICU, and free from respiratory decompensation.
7	Hospital discharge.

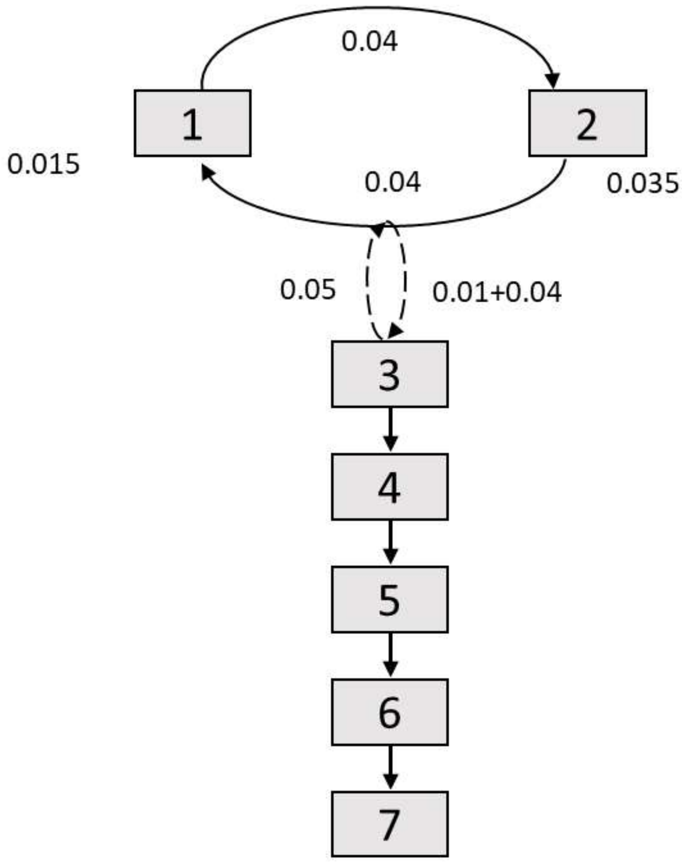
4.1.2.2 Significance level allocation for the testing procedure

The overall significance level $\alpha = 0.05$ will be split to 0.015 for H1(prevention) and 0.035 for H2 (recovery) for the initial testing. If H1 is rejected then 0.005 will be recycled and H2 will be tested at 0.035+0.005=0.04, while 0.01 will be reserved for H3. If H2 is rejected at 0.04 then H3 will be tested at full 0.05, otherwise H3 will be tested at 0.01. In the same way, if H2 is rejected at 0.035, then 0.025 will be recycled and H1 will be tested at 0.04 while 0.01 will be reserved for H3. If H1 is rejected at 0.04 then H3 will be tested at full 0.05, otherwise H3 will be tested at 0.01. If H3 is tested at full 0.05 (which can happen if and only if both H1 and H2 are rejected) and is rejected, then the testing procedure will continue for H4-H7 as a fixed sequence at full 0.05.

Implementation of this testing procedure using SAS® software is given in Appendix A 2.

As an example, consider the case of following p-values: $p_1 = 0.044, p_2 = 0.031, p_3 = 0.007$. At the initial allocation of significance level H2 will be rejected at 0.035, while H1 will not be rejected at 0.015. After recycling, H1 will be re-tested at 0.04 but will still not be rejected. H3 will be tested at 0.01, reserved from rejecting H2, and hence H3 will be rejected. Therefore, 0.01 can be recycled to test H1 at 0.04+0.01=0.05. At this level H1 will be rejected and hence the testing procedure can continue to hypotheses 4-7 at overall significance level 0.05.

Figure 2 - Multiple testing procedure



4.1.3 Presentation of Time-to-event Analyses

In general, summary tables of time-to-event analyses will include the number and percentage of patients with event per treatment group, event rate, HR with 95% confidence interval and p-value. The event rate will be derived as the number of patients with event divided by the total duration of follow-up across all patients in a given group, presented as patients with event per 100 patient months (1 month = 30 days).

Kaplan-Meier (KM) estimates of the cumulative proportion of patients with events will be calculated and plotted per treatment group, with the number of patients at risk indicated below the plot at specific time points. The KM plots will be presented for all time-to-event analyses, including the individual components of the composite endpoints.

4.1.4 Vital Status and Follow-up of the Prevention Endpoint

Potential endpoints will be collected from randomization throughout the study until and including the patient's last contact. The investigator will attempt to collect vital status (dead or alive) on or after Day 30, including vital status from publicly available sources for patients who have withdrawn consent, in compliance with local privacy laws/practices.

Known vital status at the end of the study will be defined when the patient is dead or has date last known alive on or after Day 30. In patient disposition, the number of patients who are dead, alive, or with unknown vital status will be reported separately for patients who did/did not withdraw consent. The term LTFU will be limited to only patients with unknown vital status.

The number and percentage of patients who are not LTFU will be provided, as well as the number and percentage of patients with complete follow-up of the prevention composite endpoint. In addition, the proportion of patient-days with complete follow-up will be reported per treatment group. Patient-days with complete follow-up will be defined as days from randomization until the earliest of an event of the prevention composite endpoint, WoC, censoring due to incomplete event assessment (in cases where last complete event assessment is prior to Day 30), or Day 30. To calculate the proportion of patient-days with complete follow-up, the total patient-days with complete follow-up will be divided by the total patient-days with maximum follow-up, that is the total days from randomization to first event or Day 30.

4.2 Analysis Methods

4.2.1 Demographics and Baseline Characteristics

Demographic and baseline characteristics, including medical history, will be summarized, using frequency distributions and summary statistics based on the FAS, for each treatment group as well as for all patients combined. No statistical test will be performed for comparison of any baseline measurement among treatment groups.

4.2.2 Prior, Baseline and Concomitant, Medications

Prior medication is defined as medications started before index hospitalization.

Baseline medication is defined as medications with at least one dose taken on or before date of randomization, during the index hospitalization.

Concomitant medication is defined as medications taken post randomization, irrespective of study drug.

The frequency of baseline and concomitant medication will be presented for the FAS per Anatomical Therapeutic Chemical (ATC) class and treatment group.

COVID-19 related concomitant medications will be summarized separately.

Summaries of prohibited medication (as defined in CSP Section 7.7) will be presented. In this study prohibited medication is limited to open-label SGLT2 inhibitors taken in combination with investigational product.

4.2.3 Analysis of Time-to-Event endpoints

For analysis of time to first event, data will be expressed as 2 variables:

- A binary variable indicating whether the event in question occurred or the patient was censored
- An integer variable for the number of days from randomization to the first occurrence of an event (start date of the event – randomization date + 1) or, for event-free patients, from randomization to censoring (censoring date – randomization date + 1)

Event-free patients will be censored as described below for each respective endpoint.

4.2.4 Analysis of the Prevention Primary Composite Endpoint

In the analysis of the prevention composite endpoint, treatments (dapagliflozin versus placebo) will be compared using a Cox proportional hazards model with a factor for treatment group, stratified by country (pooling together countries contributing less than 5% of all randomized patients) and adjusting for age and sex. The stratified Cox proportional hazards model is assuming proportionality of hazards in each stratum, meaning that baseline hazards in different strata can be different (which would be indicative of different event risks in different countries, during the time), but the effect of covariates over time and across strata is constant. The analysis will use the earliest of WoC, last clinical event assessment, and Day 30 for censoring of patients without any primary event as described in Section 3.1. The Efron method for ties and p-value based on the Wald statistic will be used. The event rates, p-value, HR, and 95% confidence interval will be reported.

Kaplan-Meier estimates of the cumulative proportion of patients with event will be calculated and plotted, for the composite endpoint and for the individual components.

4.2.4.1 Subgroup Analysis of the Prevention Composite Endpoint

Exploratory subgroup analyses of the primary composite endpoint will be performed for the characteristics listed in Table 3. Analysis will be done separately by each level of the relevant subgroup variable, in a model stratified by (pooled) country and including age and sex as covariates. Furthermore, in a model across subgroups a test of interaction between randomized treatment group and the subgroup variable will be performed in a Cox model, stratified by (pooled) country and including as covariates age, sex, the relevant subgroup variable and the interaction between treatment and the subgroup variable. In addition to the number and percentage of patients with event, event rate estimate, HR with 95% confidence interval and p-value for each subgroup, the interaction p-value will be presented. HRs with confidence intervals will be presented in a forest plot, also including the event rates and interaction p-values. The p-values for the subgroup analyses and interaction will not be adjusted for multiple comparisons as the tests are exploratory and will be interpreted descriptively.

Table 3 - Characteristics and categories for subgroup analysis of the primary endpoint

Characteristic	**Categories**
Age (years)[a]	<60, ≥60
Sex	Male, Female
Race	White, Black or African, Asian, Other
Country [b]	US, and other countries in Europe, Latin America and Asia
History of hypertension	Yes, No
History of atherosclerotic cardiovascular disease	Yes, No
History of HF	Yes, No
Baseline eGFR (ml/min/1.73m^2)	<60, ≥60
History of Type 2 diabetes mellitus	Yes, No
History of COPD	Yes, No
BMI (kg/m^2) at enrollment	<30, ≥30
NT-proBNP	≤median, >median
hs Troponin	≤median, >median
hsCRP	≤median, >median
D-dimer	≤median, >median

[a] Will be adjusted only for Sex and stratified by (pooled) country.

[b] The subgroup analysis by Country will exclude Country as a stratification factor from the model.

For the subgroup analysis by country countries will not be pooled.

No HR estimates with confidence intervals and p-values will be given for subgroups with less than 5 events in total, both treatment groups combined.

The following exploratory analyses will be conducted.

4.2.4.2 Supplementary Analysis of the Prevention Composite Endpoint

The prevention composite endpoint consists of 9 components:

Respiratory decompensation

1. Ventilator
2. BiPAP or CPAP

3. ECMO

Cardiac decompensation

4. New or worsened congestive HF
5. Vasopressor therapy, inotropic or mechanical circulatory support
6. Ventricular tachycardia or fibrillation, resuscitated cardiac arrest,

Renal decompensation

7. Doubling of s-Creatinine
8. Initiation of renal replacement therapy

Death

9. All-cause mortality

In reporting the events of the composite the frequency of the first events will be summarized. If a subject has events of several types on the same day, then the subject is included in the category with the highest priority (in the following order – all-cause mortality, cardiac decompensation, renal decompensation, respiratory decompensation). The contribution of each component of the prevention composite endpoint to the overall treatment effect will be examined. In the analysis of the components, all first events of the given type will be included irrespective of any preceding non-fatal composite event of a different type. Consequently, the sum of the number of patients with events in the component analysis will be larger than the number of patients with composite events. Methods similar to those described for the analysis of the composite will be used to separately analyze the time from randomization to the first occurrence of each component of the prevention composite endpoint.

In addition to analyzing the components, the following supplementary analyses will be conducted:

1. Analyzing the composite using the second definition (b) below (Supplementary #1) of the respiratory decompensation.

   a. (Primary) Respiratory decompensation requiring initiation of mechanical ventilation (includes invasive or non-invasive ventilation, CPAP, or BiPAP), and/or initiation of ECMO.

   b. (Supplementary #1) Respiratory decompensation requiring mechanical ventilation (excluding CPAP and BiPAP), and/or initiation of ECMO.

2. (Supplementary #2) Analyzing the composite by removing the respiratory decompensation component. Patients experiencing respiratory decompensation will be followed-up for other components of the prevention composite endpoint.

3. (Supplementary #3) Analyzing the composite by removing AKI (Acute Kidney Injury) component. Patients experiencing AKI will be followed-up for other components of the prevention composite endpoint.

4. (Supplementary #4) Multiple event analysis using LWYY (Lin-Wei-Lang-Ying) method (see Lin et al 2000). There are 9 components in the prevention composite endpoint, hence a patient can have at most 9 events. Only multiple events of patients will be considered, repeated events of the same component of a patient will not be counted.

4.2.4.3 Sensitivity Analysis of the Prevention Composite Endpoint

Analysis of proportions

Sensitivity analysis of the prevention composite endpoint and components will be conducted by analysis of the difference in proportions of patients with an event using the Wald test for 2-sample proportions.

Unconfirmed SARS-CoV-2 Infection

A sensitivity analysis will be performed of the primary analysis excluding patients who were not tested for SARS-CoV-2 at randomization and tested negative when testing became available.

Baseline use of remdesivir

A sensitivity analysis will be performed of the primary analysis by remdesivir use at randomization.

Incomplete in-hospital event assessment

If it is known that some patients have died after incomplete event assessment, then a sensitivity analysis will be conducted to consider these deaths as events.

Missing Data and Informative Censoring

The time-to-event analysis using the Cox regression depends on the assumption of non-informative or ignorable censoring, corresponding to the missing-at-random assumption. The missing data in this context are patients who are prematurely censored due to WoC, LTFU, or otherwise incomplete follow-up of endpoints. The amount of missing data will be described eg, in terms of the number of patients and patient time with incomplete follow-up as described in Section 4.1.4.

The amount of incomplete follow-up is expected to be small. To assess the impact of missing data and the robustness of the results with regard to the assumption of non-informative censoring, sensitivity analysis will be planned based on the evaluation of the missing follow-up and discussed in relation to the observed efficacy signal.

A tipping point analysis will be performed for patients prematurely censored before Day 30 (date of discharge for those patients that were discharged from hospital and have missing vital status assessment afterwards; the last assessment date of vital status after hospital discharge; date of withdrawal of consent for those patients that withdrew consent while in hospital). In this case, if a treatment effect is observed, we will conduct a “worst case” analysis first, where all prematurely censored patients in the dapagliflozin group will be considered as having the event at their time of censoring, while the prematurely censored patients in the placebo group will be considered as censored at Day 30. This is the most unfavorable scenario for the dapagliflozin group, since all prematurely censored patients are considered as having the event, while in placebo group these patients are considered as having longer event free time than observed. If the treatment effect is still observed, then no further analysis will be done, since this will mean that premature censoring did not affect the treatment effect. If there is no statistical significance in the worst case analysis, then events after initial censoring will be simulated in both treatment groups, using the observed event rates in each group, in the placebo group keeping the event rate constant, while gradually increasing the event rate in the dapagliflozin group, to find the event rate for which the significance is lost. Then, an increased event rate of the placebo group will be selected. The same procedure will be repeated by gradually increasing the event rate in the dapagliflozin group, to find the event rate for which the significance is lost. A shift table will be constructed for hazard ratios corresponding to various combinations of event rates in the dapagliflozin and the placebo groups.

A hazard ratio corresponding to a combination of event rates in the dapagliflozin and placebo groups will be calculated as follows: event rates per treatment group (observed or increased) will be used to generate new event times for patients who were prematurely (before Day 30) censored, using exponential distribution. If for a patient the sum of the new event time and the time spent in the study (premature censoring time) is less than or equal to 30 then it will be assumed that the corresponding patient had an event, and the sum of the observed and simulated times will be used as the time of this event. If the sum of these times is greater than 30 then this patient will be considered as censored at Day 30. After generating new event times for all patients who were prematurely (before Day 30) censored (using the previously observed events and new, simulated events), hazard ratio for the treatment effect in the dapagliflozin group will be calculated. This procedure will be repeated 100 times and corresponding hazard ratios and standard errors will be combined using the Rubin’s rule. The resulting hazard ratio will be reported in a shift table (possible values of event rates, Placebo × Dapa 10 mg). Statistically significant hazard ratios in the shift table will be marked.

4.2.5 Analysis of the Recovery Primary Endpoint (HCE)

To assess the effect of dapagliflozin on clinical status change during the 30 days of follow-up, a hierarchical composite endpoint will be considered. Each patient will be ranked based on the timing and severity of the events. Up to Day 30, all events described in the definition of the prevention composite endpoint will be considered (not just the first events) to determine the ranks of patients. Patients will be ranked using the order below:

Hierarchical composite outcome measure:

1 Time to death from any cause
2 Time to new/worsened organ dysfunction (as defined in the primary outcome measure)
   2.1 Patients experiencing more than one event of the primary composite outcome.
   2.2 Time to (only) new/worsened organ dysfunction.
3 Clinical status at Day 30 for patients still hospitalized and without any worsening organ dysfunction using the scale below
   3.1 Hospitalized, on high flow oxygen devices
   3.2 Hospitalized, requiring supplemental oxygen
   3.3 Hospitalized, not requiring supplemental oxygen
4 Time to hospital discharge

Which will result in the following ranking (from lowest to highest, where a higher rank means a better outcome), or Categories:

I. **Patients dying during the study**, [*Ranking within this cohort will be based on the timing of the event, with patients dying sooner getting a lower rank*]

II. **Patients who did not die but have more than one new or worsened organ dysfunction events**, [*Ranking within this cohort will be based on the number of events, with higher number getting a lower rank*]
A patient can have at most 8 events (there are 9 components in the prevention composite endpoint, the 9th is death which, if happened, will move the patient to category I), see Section 4.2.4.2. Only multiple events of different types will be considered, repeated events of the same component of a patient will not be counted.

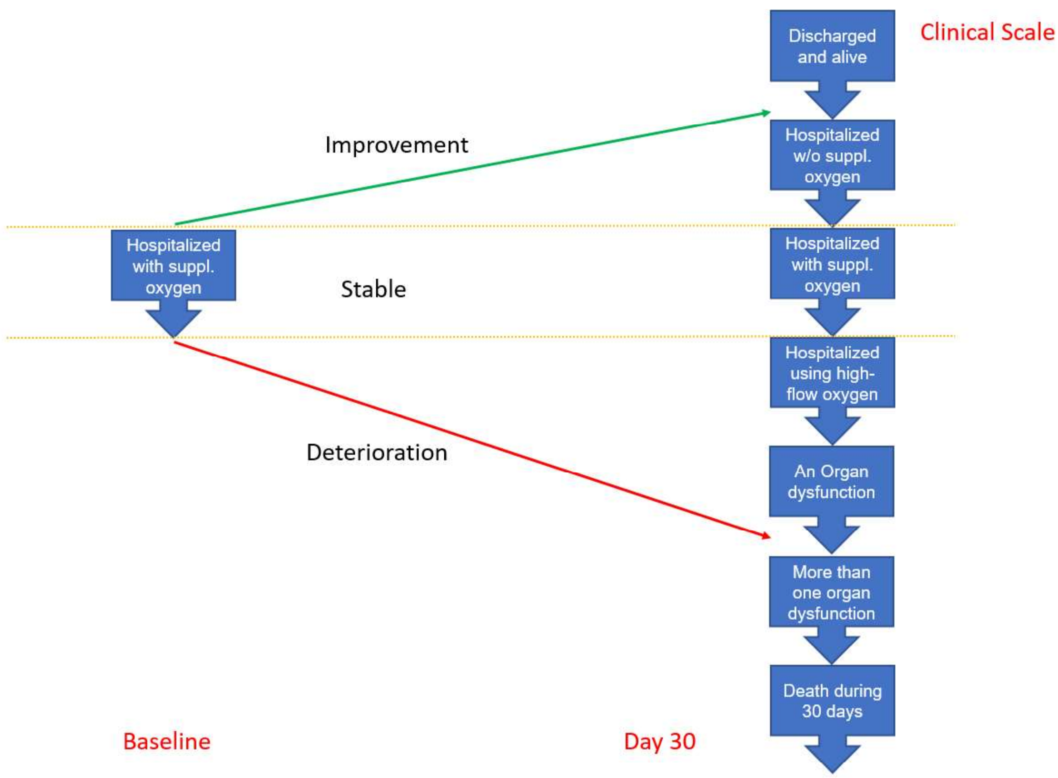
III. **Patients who did not die but have only one new or worsened organ dysfunction event,** [*Ranking within this cohort will be based on the timing of the event, with patients having the event sooner getting a lower rank. Type of organ dysfunction will not be considered*]

IV. **Patients without primary composite event but hospitalized at the end of follow-up (Day 30),** [*Ranking within this cohort, from low to high, includes patients on high-flow oxygen devices, patients requiring supplemental oxygen, and patients not requiring supplemental oxygen*]

V. **Patients alive at the end of follow up (Day 30), without primary composite event and are discharged from hospital before Day 30** will represent the highest cohort [*Ranking within this cohort will be based on the time to discharge, with patients being discharged later getting a lower rank*]

The hierarchical composite endpoint (HCE) combines clinical deterioration (e.g. organ worsening as defined in the prevention composite endpoint, prolonged hospitalization or death) with clinical improvement (e.g., change in clinical status and hospital discharge) into a single metric, and all potential intercurrent events are accounted for in this endpoint.

Figure 3 - Change in clinical status



4.2.5.1 Clinical interpretation of the recovery HCE (improvement in clinical status)

It is hypothesized that dapagliflozin will prevent in-hospital (during the index hospitalization) new/worsened organ dysfunction events and death through Day 30, which will result in more events of clinical recovery (improvement in clinical status, see Figure 3).

This hierarchical composite endpoint is similar to WHO suggested (8-point) COVID-19-specific ordinal scale recovery endpoint (WHO 2020). But there are several important differences between the ordinal scale endpoint and the suggested HCE:

1. The ordinal scale endpoints are assessed at a prespecified timepoint (for example, at Day 15) while for ranking the HCE uses severity of all events that a patient experiences during the 30 days of follow-up. For example a patient discharged from hospital before Day 15 and experienced death soon after Day 15 will be categorized as recovered in the ordinal scale, but HCE will rank this patient based on the worst experienced event, namely death.
2. The HCE takes into account in-hospital worsening of COVID-19 and not only the eventual discharge from hospital. For example, patients having in-hospital worsening before hospital discharge will be categorized as recovered in the ordinal scale, but HCE will rank this patient based on the worst experienced event, namely worsening of COVID-19.

Therefore, the HCE is a recovery endpoint with a stricter definition of recovery (discharge from hospital without a worsening event and alive, or still in hospital without a worsening event and without oxygen support, see the ”Improvement in the clinical status” line in Figure 3). Almost all patients enrolled in this study have the same baseline severity – hospitalized with COVID-19 with low-flow oxygen support. Recovery is represented on the clinical scale as improvement in clinical status compared to baseline:
1. Discharge from hospital before day 30 without in-hospital worsening and alive at Day 30; or
2. Still in hospital at Day 30, but without in-hospital worsening during the 30 days of hospitalization and without oxygen support.

Additionally, the timing of events is used to distinguish between those who recover or who experience worsening, thus making this endpoint more sensitive to capture treatment effect.

4.2.5.2 Derivation of analysis values

Based on the ranking algorithm above, for all patients analysis values will be derived as follows:

I. Patients dying during the study will have the category 400 with AVAL=400 - (study days from randomization to death).

II. Patients who did not die but have more than one new or worsened organ dysfunction events will have the category 300 with AVAL=300 + (number of events),

III. Patients who did not die but have only one new or worsened organ dysfunction event, will have the category 200 with AVAL=200 - (study days from randomization to organ dysfunction),

IV. Patients without primary composite event but who are hospitalized at the end of follow-up will have the category 100 with AVAL=100 + (1 for patients not requiring supplemental oxygen, 2 for patients requiring supplemental oxygen, and 3 for patients on high-flow oxygen devices,).

V. Patients alive at the end of follow up (Day 30), without primary composite event who are discharged from hospital before Day 30 will be represented by the category 0 with values AVAL= 0 + (study days from randomization to hospital discharge).

Note that the choice of values 100, 200, 300, 400 introduces the order between the 5 categories described above. The actual numerical values of these numbers will not affect the results since the analysis method described below is applicable for all ordinal random variables.

4.2.5.3 Handling of missing data

With the ranking algorithm described above two types of missingness are associated:

a) Events with missing date of occurrence (for example, if a patient is known to have died but the date of death is unknown).

b) Missing occurrence of events (for example patients who are lost to follow-up or withdrew consent at any time during the study).

For all analyses (primary and supplementary) type a) missingness will be handled by imputing ranks of these patients by assigning the median rank of the category to which this patient is categorized. For example, if it is known that the patient have died, then this patient is in the category I. Hence the median rank of the patients who have died can be used as an analysis value for the patient with the missing death date. Type b) missingness will be handled in the primary analysis by censoring (details below).

4.2.5.4 The primary analysis method – Stratified log-rank test

The primary analysis method for the HCE will be stratified log-rank test which will be applied to AVAL described above (see APPENDIX for implementation). The p-value from this test will be used in the multiple testing procedure (Section 4.1.2). Stratification will be done using pooled strata, similar to the analysis of the prevention composite endpoint. The treatment effect will be characterized by a win ratio (WR) estimated from the stratified Cox regression model with the Efron method for ties. The win ratio is the odds that a patient in the dapagliflozin group will have a better clinical status during the 30 days of treatment than a patient in the placebo group. The null hypothesis is that WR=1 meaning that there is no treatment effect on clinical status of patients, while WR >1 (WR < 1) would mean beneficial effect of dapagliflozin (placebo).

The advantage of calculating the win ratio estimate from Cox regression is that it allows to incorporate type b) missingness as follows.

Patients will be ranked based on the most severe event they had during the time they spent in the study using the same ranking algorithm as patients completing the 30-days of follow-up. But unlike patients completing the 30-day follow-up, patients with type b) missingness will be considered as censored in the log-rank test and Cox regression analysis.

Note that type a) missingness will be handled by imputing ranks of these patients by assigning the median rank of the category to which this patient is categorized.

Supplementary analysis – Win ratio (direct estimate)

Win ratio based on direct comparison of analysis values of two treatment groups will be calculated as a supplementary analysis. Type a) missingness will be imputed (using median rank of the category), while patients with type b) missingness will have missing analysis values, hence will not be included in the analysis.

Using the patient-level analysis values (AVAL described above), every patient in the dapagliflozin group will be compared with every patient in the placebo group. Each comparison will result in a "win", "loss" or "tie" for the patient in the dapagliflozin group if the analysis value of the patient in the dapagliflozin group is lower, higher or equal to the analysis value of the patient in the placebo group, respectively. Within the dapagliflozin group, total number of wins will be divided by the total number of losses (ties are split evenly between wins and losses) to form the Win Ratio statistic of the dapagliflozin group against the placebo group (Pocock et al 2012). The confidence interval of the win ratio statistic will be calculated as described in Gasparyan et al 2020 (see APPENDIX).

4.2.5.5 Exploratory analyses of the recovery HCE

An exploratory subgroup analysis will be conducted for the HCE using the same subgroups as in the subgroup analysis for the prevention composite endpoint (Table 3).

The following exploratory analyses will be conducted.

Supplementary analysis - different definition of respiratory decompensation

The respiratory decompensation will be redefined as requiring initiation of mechanical ventilation (excluding CPAP and BiPAP), and/or initiation of ECMO.

Sensitivity analyses

Unconfirmed SARS-CoV-2 Infection

A sensitivity analysis will be performed of the primary analysis excluding patients who were not tested for SARS-CoV-2 at randomization and tested negative when testing became available.

Baseline use of remdesivir

A sensitivity analysis will be performed of the primary analysis by use of remdesivir at randomization.

Imputing ranks based on the most severe event

In this analysis patients having type b) missingness will be imputed using their most severe event while in the study. For example, patients who are lost to follow-up cannot be categorized. But if they had an organ worsening before being lost to follow-up, then they can be categorized to category III for the sensitivity analysis. This analysis extends the follow-up time of the prematurely discontinued patients to Day 30 as it assumes that patients who prematurely discontinued the study would not have experienced more severe events than they had while in the study, if they had stayed in the study up to Day 30. Thus, in a sense it uses the last (most severe) observation. In deriving the win ratio, these patients, unlike the primary analysis, will not be censored but will contribute their rank as patients with complete follow-up. The direct win ratio will be calculated as well and imputed ranks will be used in pairwise comparisons.

Imputing ranks based on the "one-less" category

Another sensitivity analysis will be performed for patients with type b) missingness by categorizing these patients to a "one-less" category (compared to their most severe event while in the study). For example, patients who had an organ worsening before being lost to follow-up, then they can be categorized to category II for this sensitivity analysis (instead of category III). This analysis assumes that patients who prematurely discontinued the study would have experienced a more (by one category) severe event than the most severe event they had experienced while in the study.

4.2.6 Analysis of the Secondary Efficacy Variables

4.2.6.1 Time-to-Event Endpoints

The time-to-event secondary variables included in the testing hierarchy are the following:

- Time to hospital discharge.
- Time to composite kidney endpoint.
- Time to death from any cause.

These endpoints will be analyzed in the same manner as the prevention composite endpoint. Analysis of the composite of kidney endpoint, in addition to stratification by (pooled) country will be adjusted for baseline eGFR.

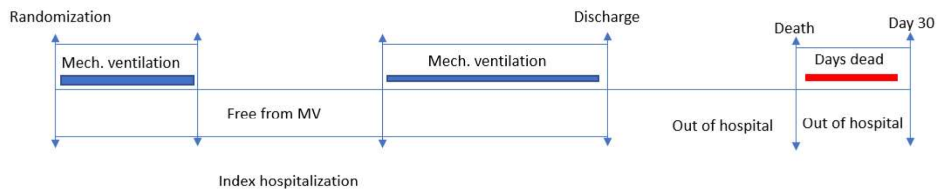
Additionally, an exploratory subgroup analysis (Table 3) will be conducted for the analysis of time to death from any cause, in the same way as the subgroup analysis for the prevention composite endpoint.

Sensitivity analyses for the time-to-event endpoints will be conducted to analyze the difference in proportions of patients with an event using the Wald test for 2-sample proportions.

4.2.6.2 Analysis of Total Number of Days

The endpoints based on total number of days (Sections 3.3.2 to 3.3.3) are similar to the endpoint "Days alive and out of hospital", introduced in Ariti et al 2011 for patients with Heart Failure. The purpose of this endpoint is to characterize the whole burden of the disease by accounting for repeated events as well as lengths of events (e.g. duration of mechanical ventilation). Higher number is indicative of a more favorable outcome since it shows more days free from mechanical ventilation. Number of days dead is subtracted from total days in order to account for death being a competing event. Subtracting the days dead from the total follow-up time has the same meaning as combining the events of mechanical ventilation and death into a composite. The total days of using mechanical ventilator when added to the total days dead will form the duration of the "negative" outcome. This value is subtracted from 30 days to obtain the duration of "positive" outcome, which is the value of interest. The four figures below show the calculation method of number of days alive and free from mechanical ventilation for an individual patient, depending on the presence of death and the type of censoring. The time interval of being out of hospital and alive is considered as being free from mechanical ventilation.

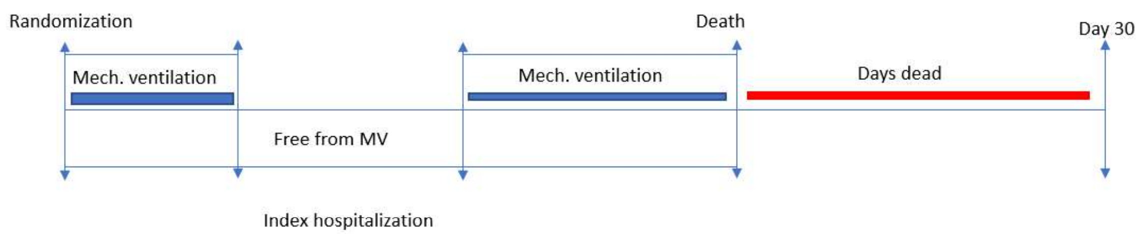
Figure 4 - Total number of days alive and free from MV – outpatient death



30 Days follow-up= (In hospital) + (out of hospital) = (MV + free of MV) + (out of hospital and alive + days dead)

# of days alive, free from MV= 30 – MV – days dead

Figure 5 - Total number of days alive and free from MV – death in hospital



30 Days follow-up= (In hospital) + (out of hospital) = (MV + free of MV) + (days dead)

# of days alive, free from MV = 30 – MV – days dead

In the cases when patients withdraw their consent or are LTFU (presented in Figure 6 and Figure 7, respectively), the follow-up time for these patients will be defined as the timeframe from randomization to the time of LTFU or WoC. In other words, if a patient withdraws from the study or is LTFU, the total follow-up time for this patient will be the time spent in the study. Then, from the total follow-up time, as above, the periods of hospitalization with use of mechanical ventilation and days dead will be subtracted.

Figure 6 - Total number of days alive and free from MV - WoC

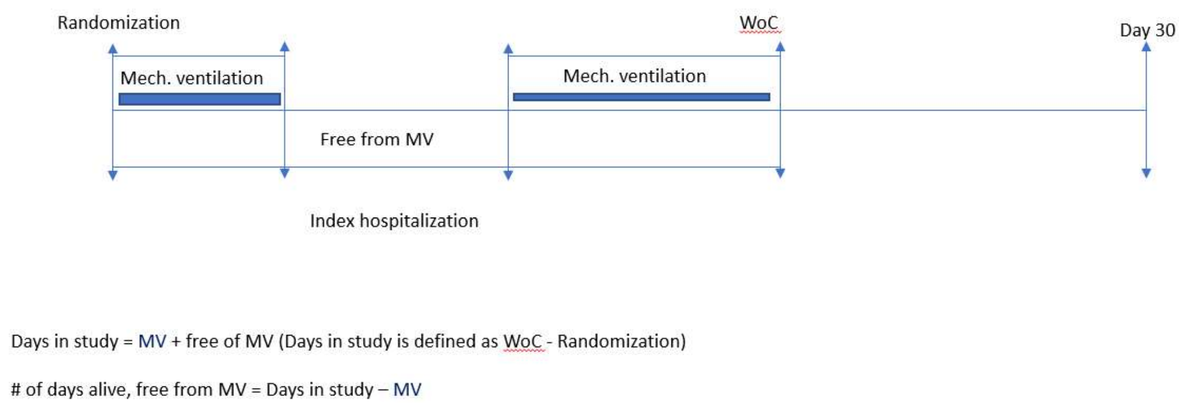
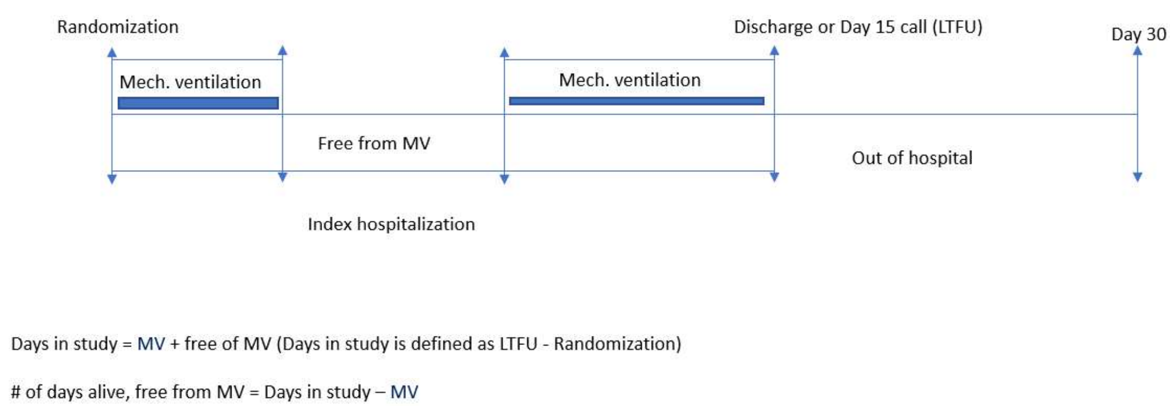


Figure 7 - Total number of days alive and free from MV - LTFU



Total number of days endpoints take the values 0,1,2,…,30, hence have discrete distributions with non-negative values.

The primary analysis will be conducted where patients first will be compared based on the vital status at the end of 30-day follow-up period and then, within each category, based on the actual values of total number of days of each patient. The analysis values (AVAL) will be derived as follows:

I. Patients who died during the 30-days of follow-up will be in the first group and will get the group rank 100. Within this group patients will be ranked based on the total number of days alive and free from MV derived above. That is, if X is the total number of days alive and free from MV for a subject who died during the 30-days of follow-up period, then for this subject AVAL=100+X.

II. Patients who do not die during the 30-days of follow-up will be in the second group and will get the group rank 200. Within this group patients will be ranked based on the total number of days free from MV derived above. For these subjects AVAL=200+X, where X is the total number of days free from MV.

Note that the assignment of values 100 and 200 introduces an order and their actual numeric values will not impact the analysis results, since for the win ratio analysis only ordinal (related to comparison) properties of variables are used. Using derived patient level analysis values, patients in two treatment groups will be compared and the patient with a higher analysis value will "win". Because of a higher group rank of patients who are alive at Day 30, patients who die during the 30-day follow-up period cannot win in comparisons against patients who are alive at Day 30. Patients with missing vital status or who withdraw consent will be considered as censored and their total number of days X (calculated in Figure 6, Figure 7) will be used to derive the AVAL using the rule of group II.

To analyze this hierarchical endpoint, the win ratio (WR) approach will be used (patients will be compared using the AVAL, and the patient with higher AVAL will "win"). The win ratio, its 95% confidence interval and the 2-sided p-value will be calculated from stratified Cox regression as in the analysis of primary recovery endpoint (see Section 4.2.5).

The numeric values of total number of days (which takes the values 0,1,…,30) will be analyzed using direct win ratio approach as a supplementary analysis, as was suggested by Wang and Pocock 2016, while the 95% confidence interval for the WR and the corresponding p-value will be calculated using the theory of U-statistics (Gasparyan et al 2020, Koch et al 1998, see APPENDIX). In this analysis patients who are lost-to-follow-up or withdrew consent will contribute as well (using their time spent in the study, see Figure 6, Figure 7). The difference from the primary analysis is that no ranking will be done based on the vital status. Therefore, in this analysis the endpoint will be a (non-hierarchical) composite endpoint where death during the study will have the same contribution as having mechanical ventilation at that point which lasts until the end of 30-days period. This will be supplementary to the hierarchical analysis and will have the advantage of providing numeric analysis values, which can be analyzed descriptively (using mean and standard deviation, for example). Number of patients with "total number of days alive and free from mechanical ventilation"=30 (patients who were alive at the end of 30-day treatment period and did not use mechanical ventilation during the study) will be summarized by treatment group.

The endpoint "Total Number of Days Alive, Not in the ICU, and Free from Mechanical Ventilation" will be analyzed in the same manner, only time interval in the ICU will be subtracted from the total number of days alive and free from mechanical ventilation.

4.2.7 Analysis of Safety Variables

Analysis Set

For safety analyses, all summaries will be based on the SAS (Section 2.1.2).

Exposure

The total exposure to study drug will be defined as the length of period on study drug, calculated for each patient as date of last dose – date of first dose +1.

An alternative measure where days of interruption are removed will be calculated and termed actual exposure. Study medication interruption is defined as a temporary medication discontinuation with an intent to restart at a later time.

Total and actual exposure will be presented descriptively.

Treatment Periods

The summaries for the on-treatment period will include events with an onset date on or after first dose of randomized study drug and on or before 2 days after last dose of study drug, up to 30 days of follow-up. Additional presentations will include all events with onset on or after first dose of study drug and up to 30 days of follow-up, regardless of whether patients are on or off study treatment at the time of the event (the "on+off" treatment period).

4.2.7.1 Adverse Events

Reportable SAEs (defined in CSP Section 6.4.2) and safety events of acute kidney injury and DKA will be recorded. Additionally, DAEs and events of "Acute renal failure", defined by narrow SMQ (Standardized MedDRA Queries) scope, will be summarized. Summaries of AEs will be limited to these categories.

AEs will be classified according to Medical Dictionary for Regulatory Activities (MedDRA), using the most current version of MedDRA possible.

Summaries by system organ class (SOC) and preferred term (PT) will be sorted by international order for SOC and by descending order of PT in the dapagliflozin treatment group.

No statistical tests to compare crude AE frequencies between treatment groups will be performed.

A summary table will be provided of the total number and percentage of patients with an SAE.

4.2.7.2 Serious Adverse Events

The number and percentage of patients with SAEs will be presented by SOC, PT, and treatment group. The most common SAEs will also be presented by PT and treatment group only.

AEs with outcome of death will be presented separately by SOC and PT.

4.2.7.3 Safety Events

Each category of safety event will be presented separately: acute kidney injury and DKA.

4.2.7.4 Laboratory Evaluation

All summaries of clinical chemistry/hematology parameters will be presented in both conventional and SI units. Study mandated laboratory values and standard of care laboratory values will be analyzed separately as well as in overall assessment.

The result and the change from baseline of each clinical chemistry/hematology tests, including estimated GFR , will be summarized by treatment group at each scheduled visit using descriptive statistics, including n, mean, SD, median and quartiles.

4.2.7.5 Marked Laboratory Abnormalities

The number and percent of patients with a marked abnormality in clinical laboratory tests will be summarized over time by treatment group.

Laboratory abnormalities will be evaluated based on marked abnormality (MA) criteria. The list of MAs is provided in Table 1 below.

An on-treatment value will be considered an MA if either

- the on-treatment value is beyond an MA limit AND the baseline value is not beyond the same limit,

  OR

- both the baseline and on-treatment value are beyond the same MA limit AND the on-treatment values is more extreme (farther from the limit) than was the baseline.

Laboratory MAs occurring during the on-treatment period will be summarized by treatment group. The directions of changes (high or low) in MAs will be indicated in the tables.

Table 4 - Marked abnormality criteria for safety laboratory variables

		Marked Abnormality Criteria	
Clinical laboratory variables	**Units**	**Low**	**High**
Hematology			
Hematocrit		< 0.20	> 0.55
Hematocrit			> 0.60
Hemoglobin	g/L	< 60 g/L	> 180 g/L
Hemoglobin	g/L		> 200 g/L
Na (Sodium)	mmol/L	< 130 mmol/L	> 150 mmol/L
Na (Sodium)	mmol/L	< 120 mmol/L	
K (Potassium)	mmol/L	≤ 2.5 mmol/L	≥ 6.0 mmol/L
Creatinine	µmol/L		≥1.5X BL CREAT
Creatinine	µmol/L		≥2X BL CREAT

BL is the baseline measurement

4.2.8 Analysis of Exploratory Objectives

The analysis of the exploratory variables will in the same fashion as the primary and secondary efficacy variables be based on the ITT principle, including data irrespective of whether the patient has discontinued study drug.

The exploratory endpoints are

- Change in NT-proBNP, hs troponin, D-dimer, LDH, ALT, lymphocyte count, CRP between Day 1 and Day 15 (or discharge from hospital, whichever is earlier)
- Qualitative PCR for SARS-CoV-2 in oropharyngeal/nasopharyngeal swab at baseline (while hospitalized); and Day 15 (if still hospitalized) or discharge from hospital.
- Change in NEWS 2 from Day 1 to Day 15 (or discharge from hospital, whichever is earlier).
- Patient’s clinical status (on a 7-point ordinal scale) at Day 15 (or discharge from hospital, whichever is earlier)
- Total number of days alive, out of hospital, and not on renal replacement therapy (during index hospitalization only)

- Proportion of patients with ACS (acute coronary syndrome).
  Acute coronary syndrome is defined as: during index hospitalization, abnormal troponin level above 99th percentile of the local laboratory reference range or, if abnormal at baseline, further rise in troponin levels accompanied by at least 1 of the following: 1) ischemic symptoms 2) ischemic ST-segment changes on ECG. (Thygesen et al 2018)

Proportions will be analyzed using 2-sample proportions tests. Endpoints with continuous values (like laboratory values) will be analyzed using generalized linear models with appropriate distributions, including fixed effects for treatment group and baseline values as a covariate. Endpoints with ordinal values will be analyzed using the WR approach.

The total number of days endpoint will be analyzed similarly to the corresponding secondary endpoint (Section 4.2.6.2).

5 INTERIM ANALYSES

Interim analysis for safety only will be performed by the Independent Data and Safety Monitoring Committee after the first 100 patients have completed the 30-day treatment period; no interim efficacy analyses are planned.

6 OBSERVATIONAL PERIOD

All efficacy and safety objectives described in the CSP are based on the blinded and treated 30-days of follow-up. As soon as the pre-planned number of patients has completed the 30-day treatment period, and the data is collected and cleaned, the database will be locked and unblinded for the analysis of 30-day treatment period data. The primary clinical study report will be based on the 30-days of follow-up data.

Following last dose of investigational product on Day 30, an extended follow-up period of an additional 60 days of observational follow-up (on top of the current active treatment duration of 30 days) will be conducted with telephone visits on Day 60 and Day 90, to examine longer-term trajectory of recovery from COVID-19 among trial participants. The following assessments will be completed:

- Concomitant medications will be recorded
- Follow up of ongoing SAEs. Details of any new SAEs will be recorded.
- Vital status
- Details of any re-hospitalization for the patient
- Patient Clinical Status will be assessed using a 7-point scale, defined as

1. Not hospitalized, no limitations on activities
2. Not hospitalized, limitation on activities
3. Hospitalized, not requiring supplemental oxygen
4. Hospitalized, requiring supplemental oxygen
5. Hospitalized, on high flow oxygen devices
6. Hospitalized, on invasive mechanical ventilation or ECMO
7. Death

Observational period will be analyzed descriptively. SAEs by System organ class/Preferred term, hospitalizations, concomitant medications and clinical status on the 7-point scale reported during the phone calls on Day 60 and Day 90 will be summarized. Listings of patient level data on resolution of SAEs emerged during the 30-day treatment period, as well as clinical status change for patients still in hospital at the end of the 30-day treatment period will be provided.

7 REFERENCES

Ariti et al 2011
Ariti CA, Cleland JG, Pocock SJ, Pfeffer MA, Swedberg K, Granger CB, et al. Days alive and out of hospital and the patient journey in patients with heart failure: insights from the candesartan in heart failure: assessment of reduction in mortality and morbidity (CHARM) program. AM Heart J 2011; 162(5), 900-906.

Burman et al 2009
Burman C-F, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Statistics in medicine 28.5, 739-761, 2009.

Cao et al 2020
Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. N Engl J Med. 2020. doi: 10.1056/NEJMoa2001282.

Dmitrienko et al 2006
Dmitrienko A, Wiens B, Westfall P. Fallback tests in dose-response clinical trials. Journal of biopharmaceutical statistics, 16(5), 745-755, 2006.

Dodd et al. 2011
Dodd LE, Follmann D, Wang J, Koenig F, Korn LL, Schoergenhofer C, Proschan M, Hunsberger S, Bonnett T, Makowski M, Belhadi D. Endpoints for randomized controlled clinical trials for COVID-19 treatments. Clinical Trials. 2020:1740774520939938.

Gasparyan et al 2020
Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. Adjusted Win Ratio with Stratification: Calculation Methods and Interpretation. Stat Methods Med Res, DOI: 10.1177/0962280220942558.

ICH E9(R1) 2019
ICH E9(R1). Harmonised Guideline Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Final version. Adopted on 20 November 2019.

FDA 2017
FDA Guidance document. Multiple Endpoints in Clinical Trials Guidance for Industry. January 2017. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry

Kawaguchi et al 2011
Kawaguchi A, Koch GG, Wang X. Stratified multivariate Mann–Whitney estimators for the comparison of two treatments with randomization based covariance adjustment. Statistics in Biopharmaceutical Research. 2011;3(2):217-31.

Koch et al 1998
Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statist. Med. 1998;17(15-16),1863-92.

Levey at al 2009
Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009 May 5;150(9):604-12.

Lin et al 2000
Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2000; 62(4):711–730.

McMurray et al 2007
McMurray JJ, Teerlink JR, Cotter G, Bourge RC, Cleland JG, Jondeau G, et al; VERITAS Investigators. Effects of tezosentan on symptoms and clinical outcomes in patients with acute heart failure: the VERITAS randomized controlled trials. JAMA.
2007 Nov 7;298(17):2009-19.

Pocock et al 2012
Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European heart journal 2012, 33(2), 176-182.

Thygesen et al 2018
Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth Universal definition of Myocardial Infarction. Circulation. 2018;138(20):e618–51.
doi: 10.1161/CIR.0000000000000617.

Wang and Pocock 2016
Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharmaceut. Statist. 2016;15:238–45.

WHO 2020
WHO, R&D. 2020. WHO R&D blueprint novel coronavirus (COVID-19) therapeutic trial synopsis. https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf

8 APPENDIX

A 1 Confidence Interval for the win ratio

Confidence intervals for the win ratio will be calculated using the formulas from Section 2.2 (Theorem 2) from Gasparyan et al 2020. Suppose that we have derived analysis values of the two treatment groups as described in Section 4.2.5. Below group 2 shows analysis values of the dapagliflozin group, while group 1 is for the analysis values of the placebo group (see ***Derivation of analysis values*** in Section 4.2.5)

$$Y_1 = (y_{11}, \ldots, y_{1n_1}), \quad Y_2 = (y_{21}, \ldots, y_{2n_2})$$

There are $n_1 n_2$ possibilities of comparing a component of Y_2 to a component of Y_1: For each comparison we can have three results—a "win" for the active group if $y_{2j} < y_{1i}$ (the analysis value of the active group is *less* than that of the placebo group) a "loss" if $y_{2j} > y_{1i}$, or a "tie" if $y_{1i} = y_{2j}$. Total number of wins of the active group plus half of ties divided by the total number of comparisons is called the Win Proportion (WP) of the active treatment against the placebo group. The win ratio of the active group against the placebo is the odds of the winning, that is WR=WP/(1-WP). While the win proportion can be calculated using the formulas below

$$\hat{\theta}_N = \frac{1}{n_2}\sum_{j=1}^{n_2}\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\mathbb{1}\{y_{2j} > y_{1i}\} + 0.5\,\mathbb{1}\{y_{1i} = y_{2j}\}\right) = \frac{1}{n_2}\sum_{j=1}^{n_2} p_j$$

here $\mathbb{1}$ is an indicator taking the value 1 if the corresponding specification is satisfied, or 0 otherwise. The quantities p_j are defined as

$$p_j = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\mathbb{1}\{y_{2j} > y_{1i}\} + 0.5\,\mathbb{1}\{y_{1i} = y_{2j}\}\right), \quad j = 1, \ldots, n_2$$

and called the individual win proportions of subject j in the active group against the placebo group. In the same way, we can define the win proportion of an individual i in the placebo group against the active group as

$$q_i = \frac{1}{n_2} \sum_{j=1}^{n_2} \left(\mathbb{1}\{y_{2j} < y_{1i}\} + 0.5\mathbb{1}\{y_{1i} = y_{2j}\} \right), \quad i = 1, \ldots, n_1$$

A confidence interval for the win proportion can be calculated from the asymptotic normality of the statistic

$$Z_N = \frac{\hat{\theta}_N - \theta}{\sqrt{\frac{var(Y_2^0)}{n_2} + \frac{var(Y_1^0)}{n_1}}}$$

here

$$var(Y_2^0) = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (p_j - \hat{\theta}_N)^2, \quad var(Y_1^0) = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (q_i - (1 - \hat{\theta}_N))^2$$

Then, applying the monotone transformation f(x)=x/(1-x) to the confidence interval of the win proportion, we will get a confidence interval and the corresponding p-value for the win ratio.

A 2 Multi-Stage Fallback testing Procedure

Denote by $p_i, \ i = 1, \cdots, 7$ the p-values of the hypotheses 1-7. The following algorithm uses two-sided p-values, therefore the treatment effect estimate should be in the direction that favors treatment as well, in order to consider the specific hypothesis rejected.

Step 1. Test the primary endpoints using the weight

$$p_1 \leq \omega\alpha, \ \ p_2 \leq (1 - \omega)\alpha.$$

If *any* of these inequalities is true, proceed to the next step.

Step 2. Then the following inequalities will be checked, where $I(\cdot)$ denotes the indicator function which takes the value 1 if the underlying inequality is true and 0, otherwise. These indicator functions show how some part of α can be recycled.

$$p_1 \leq \omega\alpha + (1 - \theta)(1 - \omega)\alpha \, \boldsymbol{I}_{(\boldsymbol{p_2 \leq (1-\omega)\alpha})}$$

$$p_2 \leq (1 - \omega)\alpha + \gamma\omega\alpha \, \boldsymbol{I}_{(\boldsymbol{p_1 \leq \omega\alpha})}$$

If *any* of these inequalities is true, proceed to the next step.

Step 3. If H1 is rejected at $\omega\alpha$, then γ part of it ($\gamma\omega\alpha$) is recycled for H2 while $(1-\gamma)$ part of it is passed to H3 ($(1-\gamma)\omega\alpha$). If H2 is rejected at $(1-\omega)\alpha$, then $(1-\theta)$ part of it is recycled for H1 ($(1-\theta)(1-\omega)\alpha$), while θ part of it is reserved for H3 ($\theta(1-\omega)\alpha$). If the initially non-rejected hypothesis is rejected at the recycled level, then the entire recycled mass is passed to H3 (hence it is tested at full α).

Therefore, the significance level for H3 is

- $(1-\gamma)\omega\alpha,$ if only H1 is rejected, that is,

$$p_1 \le \omega\alpha, \text{ but } p_2 > (1-\omega)\alpha + \gamma\omega\alpha.$$

- $\theta(1-\omega)\alpha$, if only H2 is rejected that is

$$p_1 > \omega\alpha + (1-\theta)(1-\omega)\alpha, \qquad p_2 \le (1-\omega)\alpha.$$

- α, if both H1 and H2 are rejected, that is either

$$p_1 \le \omega\alpha \ \ and \ p_2 \le (1-\omega)\alpha + \gamma\omega\alpha,$$

or

$$p_1 \le \omega\alpha + (1-\theta)(1-\omega)\alpha \ \ and \ p_2 \le (1-\omega)\alpha.$$

Step 4. It is possible to recycle the significance level if H3 is rejected, to test the remaining hypothesis H1 or H2 (whichever was not yet rejected).

$$p_1 \le \omega\alpha + (1-\theta)(1-\omega)\alpha\, \boldsymbol{I}_{(\boldsymbol{p_2 \le (1-\omega)\alpha})} + \theta(1-\omega)\alpha\, \boldsymbol{I}_{(\boldsymbol{p_2 \le (1-\omega)\alpha})}\, \boldsymbol{I}_{(\boldsymbol{p_3 \le \theta(1-\omega)\alpha})}$$

$$p_2 \le (1-\omega)\alpha + \gamma\omega\alpha\, \boldsymbol{I}_{(\boldsymbol{p_1 \le \omega\alpha})} + (1-\gamma)\omega\alpha\, \boldsymbol{I}_{(\boldsymbol{p_1 \le \omega\alpha})}\, \boldsymbol{I}_{(\boldsymbol{p_3 \le (1-\gamma)\omega\alpha})}.$$

The significance level of H3 is:

- $(1-\gamma)\omega\alpha,$ if only H1 is rejected, that is,

  $p_1 \le \omega\alpha,$ but $p_2 > (1-\omega)\alpha + \gamma\omega\alpha + (1-\gamma)\omega\alpha.$

- $\theta(1-\omega)\alpha$, if only H2 is rejected that is

$$p_1 > \omega\alpha + (1-\theta)(1-\omega)\alpha + \theta(1-\omega)\alpha \ and \ \ p_2 \le (1-\omega)\alpha.$$

- α, if both H1 and H2 are rejected, that is either

$$p_1 \le \omega\alpha \ and \ \ p_2 \le (1-\omega)\alpha + \gamma\omega\alpha \ + (1-\gamma)\omega\alpha,$$

or

$$p_1 \leq \omega\alpha + (1-\theta)(1-\omega)\alpha + \theta(1-\omega)\alpha \; and \; p_2 \leq (1-\omega)\alpha.$$

Since these steps are recursive, only the last step can be done. If and only if H3 is tested at full α and $p_3 \leq \alpha$, the testing procedure will continue to H4-H7 as a fixed sequence testing procedure.

The SAS software implementation is given below. A1-A7 show allocated significance levels for the hypotheses H1-H7. S1-S7 take value 1 if the corresponding hypothesis is rejected at the allocated level.

```
data MTP;
	alpha=0.05;
	omega=0.3;
	gamma=1/3;
	theta=2/7;

	p1=0.044;
	p2=0.031;
	p3=0.007;
	p4=0.049;
	p5=0.051;
	p6=0.01;
	p7=0.002;

	I1=p1<=omega*alpha;
	I2=p2<=(1-omega)*alpha;
	I3=p3<=(1-gamma)*omega*alpha;

	a1=omega*alpha+(1-theta)*(1-omega)*alpha*I2+theta*(1-omega)*alpha*I2*I3;
	a2=(1-omega)*alpha+gamma*omega*alpha*I1+(1-gamma)*omega*alpha*I1*I3;
	if p1<=omega*alpha and p2 > (1-omega)*alpha+gamma*omega*alpha+(1-gamma)*omega*alpha
		then a3=(1-gamma)*omega*alpha;
	else if p1 > omega*alpha+(1-theta)*(1-omega)*alpha+theta*(1-omega)*alpha
		and p2 <=(1-omega)*alpha then a3=theta*(1-omega)*alpha;
	else if (p1<=omega*alpha
	and p2<=(1-omega)*alpha+gamma*omega*alpha*(1-gamma)*omega*alpha) or
	(p1<=omega*alpha+(1-theta)*(1-omega)*alpha+theta*(1-omega)*alpha
		and p2<=(1-omega)*alpha) then a3=alpha;
	else a3=0;

	a4=(a3=alpha)*(p3<=a3)*alpha;
	a5=(a4=alpha)*(p4<=a4)*alpha;
	a6=(a5=alpha)*(p5<=a5)*alpha;
	a7=(a6=alpha)*(p6<=a6)*alpha;

	if p1<=a1 then s1=1;
	if p2<=a2 then s2=1;
	if p3<=a3 then s3=1;
	if p4<=a4 then s4=1;
	if p5<=a5 then s5=1;
	if p6<=a6 then s6=1;
	if p7<=a7 then s7=1;
run;
```

A 3 Implementation of the primary analysis for HCE

The win ratio and its 95% confidence interval from Cox regression can be calculated as follows:

```
proc phreg data=ADHCE;
	class TRTPN (ref="2");
	model AVAL*CNSR(1)=TRTPN/rl ties=efron;
	strata STRATA;
	ods output parameterestimates=est(rename=(HazardRatio=WinRatio));
run;
```

The p-value is calculated from the stratified log-rank test as follows:

```
PROC LIFETEST DATA=ADHCE notable;
	TIME AVAL*CNSR(1);
	STRATA STRATA/group=TRTP;
	ods output HomTests=LR(where=(TEST="Log-Rank"));
RUN;
```