Statistical Analysis Plan: ProSAVE

ClinicalTrials.gov Identifier: NCT04158804

Statistical Analysis Plan, Version 5.0

Jan Wiemer, Thermo Fisher Scientific, 2023-06-12

i. Signature Page

for Statistical Analysis Plan: ProSAVE

by Jan Wiemer, Thermo Fisher Scientific, 2022-04-05

This study analysis plan is approved by:

Name / Responsibility	Signature	Place, Date
Dr. Jan C Wiemer Statistician Head of Biostatistics		Hennigsdorf, 2023-06-12
Dr. Sascha Johannes Study Manager Medical Affairs Director US		Hennigsdorf, 2023-06-12
Dr. Michael Mansour Principal Investigator MGH		Boston,

ii. Version History

This statistical analysis plan (SAP) is based on the protocol version 16.0 dated Dec 20, 2022 [ProSAVE16.0].

SAP Version	Date	Change	Rationale
1	2021-12-13	Not applicable	Initial version
2	2022-01-21	Section 3.1: More precise specification of the PP population Section 4.1: More precise specification of the primary safety endpoint CAE30 concerning patients enrolled with adverse events	During the enrollment process the need arose to clarify how certain cases should be handled in the study.
3	2022-04-05	Section 4.1 & 4.6: Inclusion of the additional safety endpoint CAE30' counting events from 24h after randomization onwards Section 4.6: Inclusion of additional sensitivity analyses: comparison of results with pilot results focused on pilot study sites and overview of site specific results	Extension of the pre-specified analysis for a more comprehensive assessment of study results
4	2022-12-20	1. Order of primary endpoint and first secondary endpoint reversed 2. CAE30 non-inferiority testing instead of CAE30 superiority testing	The interim analysis with 350 patients resulted in a recommendation to stop the study due to futility. Accordingly, it is unlikely to reach CAE30 superiority at study end.

		3. Interim analyses shifted to 550 and 650 patients *4.* Corrected description how the futility criterion should be handled: stopping the study due to futility (was and) is not binding	
5	2023-06-12	*1.* Improvement of the pre-specification of the analysis of the primary endpoint (shortABx) and of the first two secondary endpoints (CAE30, prescribedABx) *2.* Extension of sensitivity analyses: pre-specification of an analysis adjusted for covariate “site” (primary endpoint and first two secondary endpoints)	1. Improve pre-specification (reduce ambiguity for strict validation) *2.* Include adjustment to covariate “site” used for stratified randomization in accordance with ICH and FDA guidelines

iii. List of Abbreviations

Abbreviation	Term
ABx	Antibiotics
AE	Adverse Events
AECOPD	Acute exacerbation of chronic obstructive pulmonary disease
AHRQ	Agency for Healthcare Quality and Research
ALT	Alanine transaminase
AST	Aspartate aminotransferase
CAE30	Composite Adverse Events until Day 30 (first secondary endpoint)
CAP	Community Acquired Pneumonia
CBC	Complete Blood Cell Count
CDC	Centers for Disease Control
CDI	Clostridium difficile infection
CRF	Case Report Form
CRO	Contract Research Organization
CRP	C-reactive protein
CURB 65	Confusion, Urea nitrogen, Respiratory rate, Blood pressure, 65 years of age and older
DOT	Days of Therapy
EC	Ethics Committee
ED	Emergency Department
GCP	Good Clinical Practice
GDH	Glutamate dehydrogenase
GMA	Global Medical Affairs
H0	Null Hypothesis
H1	Alternative Hypothesis
(HO) CDI	Healthcare facility-onset Clostridium difficile infection
IC	Informed Consent
ICF	Informed Consent Form
ICH	International Conference on Harmonization
ICU	Intensive Care Unit
IDSA	Infectious Disease Society of America
IDN	Integrated Delivery Network
IRB	Institutional Review Board
ITT	Intention to Treat (Full Analysis Set)

LOS	Length of stay
LRTI	Lower respiratory tract infection
MGH	Massachusetts General Hospital
NAAT	Nucleic acid amplification test
NI	Non-inferiority
PCT	Procalcitonin
PP	Per Protocol
prescribedABx	Antibiotics prescribed at discharge for pneumonia treatment (second secondary endpoint)
RCT	Randomized controlled trial
RR	Relative risk
SAP	Statistical Analysis Plan
shortABx	Short treatment of pneumonia with antibiotics (less than 4 days) (primary endpoint)
SOFA	Sequential organ failure assessment
qSOFA	Quick sequential organ failure assessment

1 Introduction

1.1 Background and Rational

From [ProSAVE16.0], page 20 (please refer study protocol for referenced documents):
"Antimicrobial resistance is one of our most serious health threats. Infections from resistant bacteria are now too common, and some pathogens have even become resistant to multiple types or classes of antibiotics (antimicrobials used to treat bacterial infections).[1,2] The binary issue is that when antibiotics are prescribed appropriately to treat bacterial infection**,** they are effective and should be prescribed without delay. However, 50% of the time antibiotics are misused or prescribed without proper indication.

Lower respiratory tract infections (LRTI) are among the most frequent indications for antibiotics and pose a significant risk for excessive exposure to antibiotics and increase incidence of CDI. Delay or lack of pathogen identification and non-specific clinical and radiographic findings often leave clinicians with insufficient evidence to make definitive decisions regarding the need for antibiotics. The Infectious Disease Society of America (IDSA) recommends improving antibiotic prescribing practices as an essential action. [3]

Several novel biomarkers have been proposed as a complementary strategy to evaluate the severity of bacterial infection, differentiate viral from bacterial etiologies, and improve antibiotic therapy decisions. Procalcitonin (PCT), a pro-inflammatory cytokine that is nearly undetectable in health and up regulated in response to endotoxemia, offers significantly more sensitive and specific prediction of bacterial infection. Trials comparing PCT-guided antibiotic algorithms to standard management show a significant reduction in antibiotic exposure without an increase in mortality or treatment failure.[4,5,6]

Despite this strong evidence from multiple studies a recent prospective multi-centric interventional trial in the US fell short of demonstrating antibiotic reductions by PCT-guided antibiotic management.[7] Amongst other limitations the authors of that study concluded that successful implementation of PCT may require closer educational oversight. As such, this study will compare effectiveness and safety of antibiotic prescription guided by a PCT-algorithm via a Stewardship Team over standard guidelines in hospitalized adult patients with suspected or confirmed pneumonia."

1.2 Primary Objective

The study goal is to validate that a specified PCT algorithm applied to non-COVID-19 patients hospitalized with suspected pneumonia

1. reduces antibiotic exposure specified as increase of *short* antibiotic pneumonia treatments lasting less than 4 days,
2. is non-inferior concerning safety according to composite adverse events until day 30, and
3. reduces antibiotic exposure specified as decrease of antibiotics prescribed at discharge for pneumonia treatment.

2 Study Methods and Statistical Principles

2.1 Study Design

The study is conducted according to a prospective, multi-center, randomized-controlled Bayesian adaptive design comparing the following two study arms: the "PCT arm" representing the new treatment versus the "Control arm" (or "SoC arm") representing the standard of care. Bayesian statistical inference is applied for between arm comparisons. The trial allows for early stopping for futility and early stopping for success at specific time points after pre-specified interim analyses.

2.2 Randomization

Patients are randomized with equal probability 1:1 to the two study arms. Randomization sequences were generated per study site and with block randomization applying varying block lengths of 2 to 4 assignments per study arm. The random sequences were generated in R with package "blockrand" version 3.5.2 and fed into the eCRF system which triggered the random assignments according to patient enrollment order.

2.3 Sample Size and Power Calculation

We plan to enroll an adequate number of patients for the study by a Bayesian adaptive design with a maximum number of 700 patients and interim analyses after the enrollment of 200 patients, 350 patients, 550 patients and 650 patients.

Sample size estimation was done in two steps:

(a) An initial rough estimate of study size, based on results from the pilot study for the three main endpoints (primary endpoint, and first and second secondary endpoints) and conducted for a Bayesian design with fixed sample size (samples with 200, 300, 400, 500, 600, and 700 patients).

Results were summarized in the study protocol Section 8.8 and are repeated in this SAP in Section 2.3.1.

(b) A final comprehensive examination of study characteristics of the adaptive Bayesian study design related to type-I error, statistical power and study size. In the context of these investigations, the decision parameters for final study success ("γ") and for stopping early for futility or out of success ("Θ^F" and "Θ^S") were set in such a way that the type-I error was controlled at the 2.5% level and, in addition, sufficient statistical power resulted with acceptable study size, see Section 2.3.2.

2.3.1 Initial Estimation

Sample size estimation was based on the observed results of the first two ProSAVE pilot-level study batches, see Table 2.1 (analysis population and endpoint specific results; "ProSAVE-1 & -2" comprising 95 (ITT) and 60 (PP) non-COVID patients not transferred to ICU within 24 hours, patients enrolled from 2020-05-28 to 2021-02-24). 5000 clinical trials were simulated per analysis setting (study endpoint, analysis population and assumed fixed sample size from 200 to 700 patients) with the distribution of events according to predictive posterior distributions (computed for the assumed sample sizes from pilot posterior probability distributions derived with uninformative conjugate priors). The corresponding statistical power estimates were determined as the proportions of study successes when applying Bayesian superiority testing to the three endpoints CAE30, shortABx and prescribedABx (one-sided at 2.5% significance level, i.e., with $\gamma = 0.975$).

We consider a sample size of about 500 patients to be adequate for the study because a statistical power of 90% was reached with 500 patients for the safety endpoint CAE30 assuming pilot ITT results, see Table 2.2.

In order to account for additional sources of variation and missing data, e.g. due to seasonal changes during the year, possible influences of a potential fourth wave of COVID-19 and the inclusion of additional study sites, we specify the study design in an adaptive way with planned interim analyses after 200, 350, 550 and 650 patients and a maximum number of 700 enrolled patients.

Table 2.1: Pilot study results for the safety endpoint "CAE30" and for the antibiotic exposure endpoints "shortABx" and "prescribedABx" for ITT and PP patient populations.

Simulation scenario			Pilot study results			
n	Endpoint	Data	PCT		Control	
			n	n_{pos}	n	n_{pos}
1	CAE30	pilot, ITT	48	3	47	10
2		pilot, PP	30	1	30	7
3	shortABx	pilot, ITT	48	14	47	5
4		pilot, PP	30	12	30	3
5	prescribedABx	pilot, ITT	47	15	46	24
6		pilot, PP	30	7	29	19

Table 2.2: Estimation of statistical power in terms of assurance probability (i.e. proportion of simulated trials reaching study success) for the three study endpoints "CAE30", "shortABx" and "prescribedABx" with fixed sample sizes from 200 to 700

patients (total; 100 to 350 patients per study arm) assuming effect sizes (estimate and uncertainty) of the "ProSAVE-1 & -2" pilot data results.

Data	# Patients		Statistical Power		
	Total	Per arm	CAE30	shortABx	prescribedABx
pilot, ITT	200	100	79%	82%	76%
	300	150	84%	86%	81%
	400	200	87%	90%	86%
	500	250	90%	92%	87%
	600	300	91%	93%	88%
	700	350	92%	94%	90%
pilot, PP	200	100	88%	94%	98%
	300	150	91%	96%	99%
	400	200	93%	97%	99%
	500	250	93%	98%	100%
	600	300	95%	98%	99%
	700	350	95%	98%	100%

2.3.2 More Precise Estimation for the Adaptive Design with CAE30 Superiority Testing

The cut-off parameters of final and interim analyses, γ, Θ^F and Θ^S, were specified by numerical simulations to assure a sufficiently controlled one-sided type-I error < 2.5% and a sufficiently high statistical power (study success probability). The simulations were conducted assuming different study-arm specific binary probabilities, of CAE30 events, occurrence of short antibiotic treatment and occurrence of discharge-prescribed antibiotic treatment, as strong as observed in the pilot study or of smaller effect size, e.g. half the effect size observed in the pilot study, see below.

The main simulation settings for power and type-I error evaluation were as follows:

1. Patient enrollment times were generated using a piece-wise uniform distribution with a rate of 25 patients per month for the first 7 months, 35 patients for month 8, 45 patients for month 9 and 65 patients per month thereafter.
2. Times to CAE30 for patients of the Control arm were generated using an exponential distribution with rate of 0.01 per month. These were censored by LoS data which were generated using an exponential distribution with rate = 0.12 (from pilot trial). For the PCT arm different relative risks (RR) relative to the Control arm were used to generate the CAE30 data.
3. Short and prescribed ABx data for the PCT arm and the Control arm were generated according to the observed rates in the pilot trial, and in an additional scenario with smaller differences between study arms then observed in the pilot trial.
4. Simulation for each scenario under the null (RR=1) and alternatives (RR=0.27, 0.4, 0.5, 0.6) were replicated 10,000 times and average type-I error rate and power were reported.

In addition, simulations were conducted for a lower CAE30 event rate of 12% (minimal possible event rate estimated by the study PI) instead of 20% (observed in the pilot study) for the Control arm to explore corresponding changes in study power.

Under the null (RR=1) various γ values were examined, see Table 2.3. $\gamma = 0.9785$ was chosen with an estimated type-I error of 2.48% < 2.5%.

Table 2.3: Fine tuning of the parameter γ specifying study success: estimated type-I errors.

γ	Type-I Error
0.9781	0.0256
0.9782	0.0256
0.9783	0.0253
0.9784	0.0251
0.9785	0.0248
0.9786	0.0248
0.9787	0.0247
0.9788	0.0245
0.9789	0.0245

Assuming the observed effects of the pilot trial (scenario RR = 0.27 in Table 2.4) the overall power of the trial with 700 enrolled patients is about 100% and the probability to stop the trial with 500 patients with success for the primary endpoints CAE30 and the first two secondary endpoints shortABx and prescribedABx is 92%.

If the true relative risk for CAE30 were 0.4, 0.5 or 0.6 instead, then the design would still achieve 99%, 93% and 76% overall statistical power, respectively, and the probabilities to stop with 500 enrolled patients would be 82%, 63% and 40%, respectively.

If the true effects of all three endpoints CAE30, shortABx and prescribedABx were weaker than observed in the pilot study (more specifically: $p_{Ctrl} = 0.20$ vs. $p_{PCT} = 0.10$ for CAE30 (ratio = 0.5), $p_{Ctrl} = 0.11$ vs. $p_{PCT} = 0.20$ for shortABx (ratio = 1.8) and $p_{Ctrl} = 0.51$ vs. $p_{PCT} = 0.41$ for prescribedABx (ratio = 0.8)), then the design would provide 93% statistical power for CAE30, 84% statistical power for shortABx and 62% statistical power for prescribedABx (corresponding to the hierarchical testing of study endpoints). The probability of stopping early with success with 500 enrolled patients would be about 19%.

Table 2.4: Adaptive design power for $\gamma = 0.9785$, Θ^F = 0.1 and Θ^S = 0.9 and non-binding futility with different alternative assumptions: according to "RR" for CAE30 and according to pilot study results for shortABx and prescribedABx.

Abbreviations: RR = relative risk for simulation of CAE30 events; type = analysis stage with 700 patients ("overall") and at the different interim analysis (IA) stages with 200, 350, 500 and 600 patients; power.cae = statistical power for CAE30 under the alternative with specified RR, overall and at each interim; power.sabx = statistical power for shortABx under the alternative with pilot study results, overall and at each interim; power.pabx = statistical power for prescribedABx under the alternative with pilot study results, overall and at each interim; fut.trig.prob = probability of meeting futility criteria

Adaptive Design Power with $\Theta_N^F = 0.1$ and $\Theta_N^E = 0.9$

Non-binding Futility

RR	type	power.cae	power.sabx	power.pabx	fut.trig.prob
0.27	overall	1	1	0.9976	0.0023
	IA.200				0.0022
	IA.350				0.0001
	IA.500	0.9218	0.9218	0.9215	0.0000
	IA.600	0.0542	0.0542	0.0542	0.0000
0.4	overall	0.9919	0.9917	0.9876	0.0188
	IA.200				0.0125
	IA.350				0.0034
	IA.500	0.819	0.819	0.8187	0.0016
	IA.600	0.1129	0.1128	0.1126	0.0013
0.5	overall	0.9348	0.9346	0.9297	0.0727
	IA.200				0.0346
	IA.350				0.0153
	IA.500	0.6324	0.6324	0.6319	0.0143
	IA.600	0.1535	0.1535	0.1534	0.0085
0.6	overall	0.756	0.7556	0.7502	0.2038
	IA.200				0.0762
	IA.350				0.0445
	IA.500	0.3967	0.3967	0.3963	0.0434
	IA.600	0.1552	0.1552	0.1551	0.0397

Simulation results provided by Cytel

Table 2.5: Adaptive design power for γ = 0.9785, Θ^F = 0.1 and Θ^S = 0.9 and non-binding futility with the additional alternative assuming that the effects of the three endpoints CAE30, shortABx and prescribedABx are weaker than observed in the pilot study; more specifically the assumptions were (a) for CAE30: p_{Ctrl} = 0.20 vs. p_{PCT} = 0.10, (b) for shortABx: p_{Ctrl} = 0.11 vs. p_{PCT} = 0.20, (c) for prescribedABx: p_{Ctrl} = 0.51 vs. p_{PCT} = 0.41.
Abbreviations: as in Table 2.4

Add. Scenario, Adaptive Power, $\Theta_N^F = 0.1$, $\Theta_N^E = 0.9$

Non-binding Futility

RR	type	power.cae	power.sabx	power.pabx	fut.trig.prob
0.5	overall	0.9304	0.8422	0.6214	0.0710
	IA.200				0.0350
	IA.350				0.0160
	IA.500	0.1892	0.1886	0.186	0.0118
	IA.600	0.1578	0.1572	0.154	0.0082

Simulation results provided by Cytel

2.3.3 Update for Protocol Changes (2022-12-20)

Additional clinical trial simulations were conducted to assess if the specified study design would still be adequate and provide reasonable statistical power when then following study protocol changes were conducted: (a) exchange of primary endpoint and first secondary endpoint (new primary endpoint: shortABx, new first secondary endpoint: CAE30), (b) change of CAE30 inference from superiority to non-inferiority, and (c) change of the times of interim analyses from 500 and 600 patients to 550 and 650 patients, respectively. The simulations were conducted analogously to the ones described in the previous subsection (clinical trial Monte Carlo simulations based on pilot study results) including some improved approximations of data generation according to current study knowledge (blinded analysis results of IA350 data in addition to pilot results: enrollment rate held constant at one patient per day, marginal endpoint rates over both study arms set to 20% for CAE30, 22% for shortABx and 40% for prescribedABx) and of protocol implementation (inclusion of five follow-up days of last enrolled patient at interim analysis according to data transfer experience). Bayesian CAE30 superiority testing

$$\text{Probability}(\text{CAE30 probability of PCT arm} < \text{CAE30 probability of Control arm}) > \gamma$$

was replaced by CAE30 non-inferiority testing with M = 10% non-inferiority margin

$$\text{Probability}(\text{CAE30 probability of PCT arm} < \text{CAE30 probability of Control arm} + M) > \gamma, \quad M = 0.1.$$

The protocol changes required a re-specification of the analysis parameter γ to assure type-I error rate control at 2.5% (new order of hierarchical testing of endpoints: first shortABx, then CAE30, then prescribedABx; null hypothesis for shortABx instead of CAE30; interim and final CAE30 NI analyses on PP populations; interim analyses with 550 and 650 patients). Under the null for the primary study endpoint shortABx (relative risk of shortABx for PCT arm versus Control arm set to 1) and under the additional conservative assumptions of effects of CAE30 and prescribedABx as in the pilot study ($RR_{CAE30} = 0.30$, $RR_{prescribedABx} = 0.60$; PCT arm versus Control arm) and the PP population as a random subset of 50% of the patients of the full population (50% estimated for the running study), a γ value of 0.985 was determined to keep the type-I error rate below 2.5% (increase of γ vs. previous protocol version from 0.9785 to 0.985; γ estimated by linear regression of type-I error rate vs. specified γ values based on Monte Carlo runs with 50,000 simulated clinical trials).

Moreover, simulations showed that adequate statistical power could be achieved even for small CAE30 effects. Assuming (a) half of the shortABx effect observed in the pilot study ($RR_{shortABx} = 1.9$), (b) only one tenth of the CAE30 effect observed in the pilot study ($RR_{CAE} = 0.9$), (c) half the prescribedABx effect observed in the pilot study ($RR_{prescribedABx} = 0.80$), and (d) the PP population as a random subset of 50% of the patients of the full population (50% PP patients estimated for the running study), we obtained 99% statistical power for shortABx superiority on ITT population, 75%

for CAE30 non-inferiority on PP population, and 48% for prescribedABx superiority on ITT population. The following Table 2.6 gives an overview of the study design characteristics of this scenario.

Table 2.6: Characteristics of the study design with protocol changes (sABx as primary endpoint, CAE30 non-inferiority testing with 10% NI margin, IA with 550 and 650 patients, and $\gamma = 0.985$) assuming (a) half the effect observed in the pilot study for shortABx ($RR_{shortABx}$ = 1.9), (b) only one tenth of the CAE30 effect observed in the pilot study (RR_{CAE} = 0.9), (c) half the effect observed in the pilot study for prescribedABx ($RR_{prescribedABx}$ = 0.80), and (d) the PP population as a random subset of 50% of the patients of the full population. Interim, final and total results in rows, first two columns to describe the level of analysis, columns three to five for the probabilities of meeting endpoint-specific study success criteria (hierarchical testing: (1) shortABx superiority on ITT population, (2) CAE30 non-inferiority on PP population, (3) prescribedABx superiority on ITT population, columns six to eight for outcomes of interim analyses.
Abbreviations: "IA" for "interim analysis", "IA-200" for "IA with 200 enrolled patients", "End-700" for "final analysis with 700 enrolled patients", "-" for "not applicable" and "recom." for "recommendation".

Level of Analysis		**Results**					
Stage of Analysis	**# PP Patients**	**Probability of Meeting EP Success Criteria**			**Interim Analysis Outcome**		
		sABx	**CAE30**	**pABx**	**Futility recom.**	**Study stopped**	**Efficacy**
IA-200	100	-	-	-	7%	-	0%
IA-350	176	-	-	-	8%	-	0%
IA-550	276	27%	27%	27%	12%	27%	27%
IA-650	324	12%	12%	12%	15%	12%	12%
End-700	350	59%	36%	9%	-	-	-
Total	-	99%	75%	48%	-	39%	39%

2.4 Framework of Statistical Inference

The main analyses of the study comprise statistical testing of the treatment "PCT arm" versus the "Control arm" concerning the following three binary endpoints: (a) the primary endpoint short antibiotic treatments (shortABx), (b) the first secondary endpoint composite adverse events "CAE30", and (c) the second secondary endpoint of antibiotic prescription at discharge (prescribedABx). See Section 4.1 for the definition of the endpoints.

The endpoints will be tested in a hierarchical manner: first, superiority testing will be conducted for the primary endpoint shortABx (see Section 4.4 ”Primary Analysis”). If superiority is concluded for shortABx, then non-inferiority testing will be conducted for the first secondary endpoint CAE30 (see Section 4.5). If non-inferiority is concluded for CAE30, then superiority testing will be conducted for the second secondary endpoint prescribedABx (see Section 4.6). This hierarchical testing scheme allows to test the three endpoints without correction for multiple testing.

A more detailed description of the analyses is given below.

2.5 Statistical Interim Analyses and Early Stopping Guidance

Interim analyses are planned with about 200 patients, 350 patients, 550 patients and 650 patients. Follow-up data will be considered according to their availability at the times when the interim analyses will be performed. A maximum number of 700 patients will be enrolled in the study. The interim analyses will be conducted to check for early stopping due to futility and/or success.

After the enrollment of 200 and 350 patients an interim analysis will be conducted applying a criterion for potentially *stopping for futility*. A recommendation to stop enrollment of patients for the study will be given to the principal investigator if interim results are in the futility zone to reach the CAE30 success criterion. Here, success refers to showing non-inferiority with non-inferiority margin M = 10%. More specifically, a recommendation to stop patient enrollment will be given if the posterior predictive probability of CAE30 success with the maximum number of 700 patients is below the futility cut-off $\Theta^F = 0.1$:

Probability(CAE30 success with 700 patients | interim data) $< \Theta^F$, i.e.
Probability(Posterior predictive probability($p_{PCT} < p_{ctrl} + M$, 700 patients | interim data) $> \gamma$) $< \Theta^F$,
with $\gamma = 0.985$ and M = 0.1 (10% non-inferiority margin).

Note that the simulations of study design characteristics, and in particular the specification of γ for type-1 error control at a one-sided level of 2.5%, were conducted with non-binding stopping for futility (i.e., simulated clinical trials were never stopped for meeting the futility criterion). Accordingly, when given study recommendations to stop patient enrollment for futility will be non-binding.

After the enrollment of 550 and 650 patients an interim analysis will be conducted applying criteria to stop potentially (a) for futility or (b) for study success:

a. ***Stopping for futility***
A recommendation to stop patient enrollment for futility will be given if interim results are in the futility zone to reach the CAE30 success criterion. More specifically, a recommendation to stop patient enrollment will be given if the posterior predictive probability of CAE30 success to show non-inferiority with the maximum number of 700 enrolled patients is below a specific futility cut-off $\Theta^F = 0.1$:

Probability(CAE30 success with 700 patients | interim data) < Θ^F, i.e.
Probability(Posterior predictive probability($p_{PCT} < p_{Ctrl} + M$, 700 patients | interim data) $> \gamma$) < Θ^F,
with $\gamma = 0.985$ and M = 0.1 (10% non-inferiority margin).

b. ***Stopping for success***
Enrollment will be stopped for success if interim results reveal a sufficiently high predictive probability that success criteria can be reached for the endpoints CAE30, shortABx and prescribedABx with the currently enrolled patients. More specifically, enrollment will be stopped if the posterior predictive probabilities of the three endpoints CAE30, shortABx and prescribedABx for the respective number of enrolled patients are above the specific success cut-off $\Theta^S = 0.9$:

Probability(CAE30 success with enrolled patients | interim data) > Θ^S AND
Probability(shortABx success with enrolled patients | interim data) > Θ^S AND
Probability(prescribedABx success with enrolled patients | interim data) > Θ^S .

Note: A final analysis will still be conducted following an early stopping decision at any interim after all enrolled patients have completed their respective follow-up of 30 days.

Note: Use of the posterior predictive probability will allow to conduct the interim analysis in a timely manner without the need to wait for the completion of the 30 days follow-up times of all patients. Thereby, missing or incomplete endpoint data of enrolled patients will be imputed by predicted endpoint data including its respective level of uncertainty.

Posterior predictive probabilities for the safety endpoint CAE30 will be calculated as follows:

1. Computation of predictive distributions:
   a. For patients still to be enrolled for a maximum of 350 + 350 study patients the following equation will be used to determine the beta-binomial predictive distribution (see e.g. [Cowles2013] p. 62, or [Rajat2021] section 3.1):

$$BBP(x'_i \mid n'_i, x_i, n_i) = \binom{n\prime_i}{x\prime_i} \frac{Beta(x'_i + x_i + 1, n'_i - x'_i + n_i - x_i + 1)}{Beta(x_i + 1, n_i - x_i + 1)} \quad (1)$$

with

i : index for study arm (“PCT arm” or “Control arm”),
n'_i : number of patients still to be enrolled for study arm *i*,
i.e. 350 – number of patients enrolled for study arm *i* until interim analysis,
x'_i : number of patients with events among the n'_i patients still to be enrolled for study arm *i*,
n_i : number of patients of study arm *i* with complete follow-up data,
x_i : number of patients of study arm *i* with complete follow-up data and observed events,
$Beta(\,,\,)$: beta function.

b. For already enrolled patients with partial follow-up data and no observed CAE30 yet, the following equation (2) will be applied with variables … depending on follow-up time to determine the beta-binomial predictive distribution (analogous to (a) but with interim data adequately conditioned on follow-up time, see e.g. [Rajat2021] section 3.2):

$$BBP\left(x'_i(t_j) \mid n'_i(t_j), x_i(t_j), n_i(t_j)\right) =$$

$$\binom{n'_i(t_j)}{x'_i(t_j)} \frac{Beta(x'_i(t_j)+x_i(t_j)+1, n'_i(t_j)-x'_i(t_j)+n_i(t_j)-x_i(t_j)+1)}{Beta(x_i(t_j)+1, n_i(t_j)-x_i(t_j)+1)} \quad (2)$$

with

i : index for study arm (“PCT arm” or “Control arm”),
$t = (t_1, t_2, \dots, t_p)$: vector of *p* distinct partial follow-up times (in days, $t_p < 30$),
$n'_i(t_j)$: number of enrolled patients of study arm *i* with partial follow-up data and without event until time t_j,
$x'_i(t_j)$: number of patients with events after time t_j among the $n'_i(t_j)$ patients of study arm *i* with partial follow-up data and without event until time t_j,
$n_i(t_j)$: number of patients with complete follow-up data and without event until time t_j,
$x_i(t_j)$: number of patients with events after time t_j among the $n_i(t_j)$ patients with complete follow-up data without event until time t_j ,
$Beta(\,,\,)$: beta function.
The main idea behind this computation is to apply the equation (1) presented in a. in order to compute time-dependent posterior probabilities and posterior predictive probabilities that depend on the follow-up times of adequately censored patients (the more follow-up time has already passed, the less likely it is to still observe an event).

2. With the predictive distributions we will conduct Monte Carlo simulations, e.g. simulate the number of adverse events per study arm for many (e.g. 1e6) trials with maximum number of 350 + 350 patients. The number of events per study arm is the sum of three components:
(i) the number of already observed events, (ii) the number of events predicted for patients yet to be enrolled and (iii) the number of events predicted for patients with partial follow-up data. The number of events according to (i) is fixed for all trials while the number of events according to (ii) and (iii) varies between trials and is sampled according to the posterior predictive distributions.

3. The posterior predictive probability of study success is estimated by the proportion of simulated trials with 350 patients enrolled per study arm meeting study success criteria with $\gamma = 0.985$.

Analogous computations will be conducted for the other two main study endpoints shortABx and prescribedABx.

2.6 Timing of Final Analysis

A maximum number of 700 patients will be enrolled. The study may be stopped earlier due to futility (non-binding criterion) or success (binding criterion) after interim analyses with about 200 patients, 350 patients, 550 patients and 650 patients (see previous section for more information). In any case the final analysis will be conducted comprising all patients enrolled into the study until the implementation of study stop. The final analysis will be conducted with complete 30 day follow-up data of all enrolled patients.

3 Study Population

Hospitalized adult patients with suspected or confirmed pneumonia at time of admission or while admitted to the ED who are prescribed antibiotics (including septic patients with respiratory focus) are enrolled into the study. Please refer to the study protocol for additional medical information, e.g. concerning eligibility criteria (inclusion and exclusion criteria in Section 4.2.1 and 4.2.2 of the protocol, respectively).

3.1 Analysis Populations

The intention-to-treat ("**ITT**") patient population will comprise all randomized non-COVID-19 patients. The per-protocol ("**PP**") patient population will comprise all ITT patients (a) fulfilling study inclusion criteria, (b) not meeting study exclusion criteria at enrollment, (c) even if meeting study exclusion criterion after enrollment or meeting a pre-defined overruling criterion, (d) with PCT measurements performed daily in the first week and thereafter at least every 48 hours during antibiotic treatment, (e) -if assigned to PCT **group, antibiotic steward** ship team recommendation given to the treating physician on enrollment day and for each PCT measurement taken Monday-Friday described under d) and antibiotics stopped according to the PCT algorithm. An exception will be made if the ASP decides the patient will be given the same recommendation throughout their hospitalization.

Consequently, it should be noted that ASP recommendations should also be provided for patients meeting pre-defined overruling criteria (e.g., need for ICU care) since such patients may still qualify for the per-protocol population.

Patients of the PCT arm who are discharged before that day's PCT has resulted will be excluded from the per protocol population. Patients in the PCT arm who were continued on antibiotics despite a weekend/holiday/no ASP member available PCT value that would have resulted in an ASP recommendation to stop antibiotics will also be excluded from the per protocol population. Patients in the PCT arm whose antibiotics were discontinued despite weekend/holiday/no ASP member available PCT value that would have resulted in continuation of antibiotics will also be excluded.

(f) if assigned to control group, no PCT measurements to treating physician during index hospitalization.

Primary analysis of the superiority endpoint "antibiotic exposure" will be conducted on the basis of the ITT patient population. Analysis of the non-inferiority endpoint "composite adverse events until

day 30" will be conducted on the basis of PP and ITT patient population. PP-results will be compared and discussed with corresponding results obtained for ITT patient population.

3.2 Baseline Patient Characteristics

See Section 4.2.

4 Statistical Analysis

4.1 Mathematical Definitions

4.1.1 Short antibiotic treatments ("shortABx", primary endpoint)

will be encoded as the proportion of patients per study arm (PCT arm, control arm) with less than 4 days of antibiotic duration

"antibiotic duration" (third secondary endpoint) is defined as the number of days per patient with antibiotic treatment for pneumonia comprising the number of days of antibiotic treatment during hospitalization ("ABx-days$_{hospital}$", starting with the first day of PCT measurement which is one day before enrollment into the study, counting all days with antibiotics not marked "not used for pneumonia") and the number of prescribed days of antibiotic treatment at hospital discharge ("ABx-days$_{prescribed\ at\ discharge}$", counting all days with antibiotics marked "to treat pneumonia"). The resulting sum will be bounded up to a maximum follow-up time of 30 days:

$$\text{Antibiotic duration} = \min(30, \text{ABx-days}_{hospital} + \text{ABx-days}_{prescribed\ at\ discharge})$$

4.1.2 Composite adverse events until day 30 ("CAE30", first secondary endpoint)

will be encoded as the patient level Boolean inclusive disjunction ("OR") of the following eight binary (true/false) single adverse event endpoints (1a) all-cause in-hospital mortality, (1b) all-cause mortality after discharge (as available), (2) septic shock (vasopressor use for > 1 h), (3) mechanical ventilation (via endotracheal tube for respiratory failure), (4) needed dialysis, (5) lung abscess/empyema/cavitation/necrotizing pneumonia or (6) hospital readmission.

In the case a patient is enrolled with mechanical ventilation, such an event is not considered to be an adverse event due to its occurrence before any intervention. If the patient is then extubated during their hospital stay, re-intubation is only classified as an adverse event if it occurs at least 24 hours post extubation.

In the case a patient is enrolled with septic shock, such an event is not considered to be an adverse event due to its occurrence before any intervention. If the patient then has a resolved septic shock during hospital stay, developing a new septic shock is only classified as an adverse event if it occurs at least 24 hours post resolution of the initial septic shock.

In the case a patient develops a safety adverse event at any point on the day of enrollment post randomization, such an event will be classified as an adverse event.[1]

The variable CAE30 mirrors the safety endpoint for pneumonia used in the study ProACT [Huang2018]. And, also as in ProACT, the endpoint will be analyzed in the form of study-arm-specific proportions, see Section 4.3ff.

4.1.3 **Antibiotic exposure at discharge** ("prescribedABx", second secondary endpoint) will be encoded as the proportion of patients per study arm (PCT arm, control arm) with antibiotic prescription for pneumonia treatment at discharge from index hospitalization.

The numerical endpoint **antibiotic duration** is the third secondary endpoint, see definition in Section 4.1.1.

4.1.4 The fourth secondary endpoint "antibiotic duration"

will be computed per patient as described in Subsection 4.1.2.

4.1.5 The fifth secondary endpoint "days of therapy per 1000 patient days"

will be computed per patient i as follows (i: index distinguishing different study patients):

DoT(patient i) = antibiotic duration(patient i) / los30(patient i) x 1000

with length of stay of patient i

los30(patient i) = min(30, number of hospital days of patient i).

Note that DoT(patient i) is not capped by 1000 because antibiotic duration also includes antibiotics days prescribed at discharge while los30 only counts the number of in-hospital days.

4.1.6 The sixth secondary endpoint "antibiotic days prescribed at discharge"

will be computed per patient as the number of prescribed days of antibiotic treatment at hospital discharge counting all days with antibiotics marked "to treat pneumonia".

4.1.7 The seventh secondary endpoint "treatment or readmission for CDI until day 30 after discharge"

will be encoded in binary Boolean form ("true": patient treated or re-hospitalized for CDI until day 30 after discharge, otherwise "false").

4.1.8 The eighth secondary endpoint "length of stay in hospital until day 30" ("los30")

[1] It is unlikely that early safety adverse events on the day of patient randomization can be prevented or induced by the specified PCT algorithm. Accordingly, to control for the impact of such early adverse events on study safety results, we will perform an additional safety analysis focused on safety adverse events that occur at least 24 hours after patient randomization (CAE30'), see Section 4.6.

will be computed per patient as described in the previous subsection.

4.1.9 The nineth secondary endpoint "length of stay in ICU" ("losICU30")

will be computed analogously to los30 but will only count days on ICU:

losICU30 = min(30, number of ICU days of patient i).

4.1.10 The tenth secondary endpoint "ICU admission until day 30"

will be encoded per patient in binary form ("true": patient admitted to ICU within study follow-up time of 30 days, otherwise "false").

4.1.11 Exploratory Endpoints

All the individual variables constituting the composite safety endpoint "composite adverse events until day 30" will be considered as exploratory endpoints and be encoded in Boolean-binary form (true/false).

4.2 Descriptive Statistics

The first step of statistical data analysis will comprise the application of conventional data cleaning and checking procedures including assessment of missing data and outliers, and uni- and bivariate plausibility checks. Data verifications and if applicable corrections will be requested.

Next, a thorough description of the overall study sample and of the two randomized study arms will be conducted for the two analysis populations ITT and PP. The study population will be described statistically with regard to main captured clinical patient characteristics, e.g. age, sex, ethnicity, disease status and PCT concentration. Frequency counts will be reported for nominal variables and means with standard deviations or medians and inter-quartile ranges will be reported for numerical variables as appropriate to variable type. Baseline patient characteristics will be reported for all patients (both study arms together) as well as per-study arm.

4.3 Primary Study Hypothesis

For the primary study endpoint "**short antibiotic treatments**" ("shortABx"; proportion of patients with short antibiotic treatments of pneumonia less than 4 days) the study hypotheses for statistical inference are as follows:

H0: The probability of patients with short antibiotic treatment of pneumonia is smaller or equal in the PCT arm versus the control arm.

H1: The probability of patients with short antibiotic treatment of pneumonia is greater in the PCT arm than in the control arm.

*Note that t*he study goal is one-sided: to validate a shortABx probability in the PCT arm that is higher than the shortABx probability in the control arm (see Section 4.4; one-sided statistical testing at 2.5% significance level controlled for multiple testing / interim analyses, see Section 2.3 and 2.5). Bayesian statistics with uninformative priors will be applied for analysis.

4.4 Primary Analysis of Antibiotic Exposure

The primary endpoint "short antibiotic treatments" will be analyzed for superiority. We will assess the differences in proportions by Bayesian analysis assuming binomially distributed events. Posterior probabilities will be computed study-arm specifically using conjugate beta-distributed and uninformative uniform priors (shape parameters $\alpha = \beta = 1$).

More specifically, the following generative model will be used:

$N^{s}_{PCT} \sim$ Binomial(N_{PCT}, p_{PCT})
$N^{s}_{Ctrl} \sim$ Binomial(N_{Ctrl}, p_{Ctrl})
priors: p_{PCT}, $p_{Ctrl} \sim$ Uniform(0, 1)

with

N^{s}_{PCT}, N^{s}_{Ctrl} : number of patients with shortABx of the PCT arm and control arm, respectively
N_{PCT}, N_{Ctrl} : number of patients randomized to the PCT arm and control arm, respectively
p_{PCT}, p_{Ctrl} : shortABx probability for patients of the PCT arm and control arm, respectively

Differences between beta-distributed study-arm specific posterior probabilities will be computed by Monte Carlo simulation with sampling from study-arm specific posterior probabilities. Success of the analysis of antibiotic exposure is specified by more than 98.5% posterior probability for more frequent short antibiotic treatments in the PCT arm than in the control arm, i.e.,

Probability[prob(short-ABx | PCT arm) > prob(short-Abx | control arm)] > 98.5%

with

prob(short-ABx | PCT arm) : probability of short antibiotic treatment for patients randomized to the PCT arm,
prob(short-ABx | control arm) : probability of short antibiotic treatment for patients randomized to the control arm.

The primary analysis will be conducted based on the ITT population. The analysis will also be conducted based on the PP patient populations and ITT- and PP-results will be compared and discussed.

4.5 Secondary Analysis of CAE30

The first secondary endpoint "composite adverse events until day 30" (CAE30) will be analyzed for safety non-inferiority of the PCT arm vs. the Control arm on the basis of the PP patient population. The analysis will also be conducted on the basis of the ITT patient populations and ITT- and PP-results will be compared and discussed. Differences in proportions of CAE30 between PCT arm and Control arm will be assessed by Bayesian analysis assuming binomially distributed events. Posterior probabilities will be computed study-arm specifically using conjugate beta-distributed and

uninformative uniform priors (shape parameters $\alpha = \beta = 1$). More specifically, the following generative model will be used:

$NE_{PCT} \sim Binomial(N_{PCT}, p_{PCT})$
$NE_{Ctrl} \sim Binomial(N_{Ctrl}, p_{Ctrl})$
priors: $p_{PCT}, p_{Ctrl} \sim Uniform(0, 1)$

with

NE_{PCT}, NE_{Ctrl} : number of patients with CAE30 of the PCT arm and Control arm, respectively
N_{PCT}, N_{Ctrl} : number of patients randomized to the PCT arm and Control arm, respectively
p_{PCT}, p_{Ctrl} : CAE30 probability for patients of the PCT arm and Control arm, respectively

Differences between beta-distributed study-arm specific CAE30 posterior probabilities will be computed by Monte Carlo simulation sampling CAE30 probabilities p_{PCT}, p_{Ctrl} from their analytically computed study-arm specific posterior probabilities. Study success is specified by more than $\gamma = 0.985$ posterior probability for a CAE30 probability of patients of the PCT arm smaller than the CAE30 probability of the Control arm plus M = 10% (NI margin):

Probability(CAE30 probability of PCT arm < CAE30 probability of Control arm + M) > γ ,

or with the above introduced notation

Probability($p_{PCT} < p_{Ctrl} + M$) > γ .

See Section 2.3 for the computation of the parameter γ which was set to $\gamma = 0.985$.

4.6 Secondary Analysis of Antibiotic Exposure

The second secondary endpoint “antibiotic exposure at discharge” will be analyzed for superiority. We will assess the differences in proportions by Bayesian analysis assuming binomially distributed events. Posterior probabilities will be computed study-arm specifically using conjugate beta-distributed and uninformative uniform priors (shape parameters $\alpha = \beta = 1$). More specifically, the same generative model will be used as the one for short antibiotic treatments (see Section 4.4). Differences between beta-distributed study-arm specific posterior probabilities will be computed by Monte Carlo simulation with sampling from study-arm specific posterior probabilities. Success of the analysis of antibiotic exposure at discharge is specified by more than $\gamma = 0.985$ posterior probability for less antibiotic exposure of the PCT arm than of the Control arm, i.e., more than $\gamma = 0.985$ probability that prescriptions of antibiotics for pneumonia are less probable for patients of the PCT arm than for patients of the Control arm:

Probability[prob(discharge-ABx | PCT arm) < prob(discharge-ABx | Control arm)] > γ

with

prob(discharge-ABx | PCT arm) : probability of prescribed antibiotics at hospital discharge for patients randomized to the PCT arm,

prob(discharge-ABx | Control arm) : probability of prescribed antibiotics at hospital discharge for patients randomized to the Control arm,

$\gamma = 0.985$.

The endpoints will be tested in a hierarchical manner: first, superiority testing will be conducted for the primary endpoint shortABx (see Section 4.4). If superiority is concluded for shortABx, then non-inferiority testing will be conducted for the first secondary endpoint CAE30 (see Section 4.5). If non-inferiority is concluded for CAE30, then superiority testing will be conducted for the second secondary endpoint prescribedABx. This hierarchical testing scheme allows to test the three endpoints without correction for multiple testing.

4.7 Sensitivity Analysis

Sensitivity analysis will be conducted to assess the robustness of primary and secondary analysis results.

- The results obtained for ITT and PP patient population will be compared and discussed.
- Differences between study-arm-specific posterior probabilities will also be determined by the following Bayesian MCMC analysis adjusting effects between study arms for covariate "site" (used for stratifying randomization). The primary study endpoint (shortABx) and the first two secondary study endpoints (CAE30, prescribedABx) will be assumed to be distributed binomially

  $$\text{endpoint} \sim \text{Binomial}(1, p),$$

  with probability "p" parameterized by a global intercept "a", site-specific intercepts "b_{site}" and study-arm specific intercepts "c_{arm}"

  $$\text{logit}(p) = a + b_{site} + c_{arm},$$

  using uninformative priors

  $$a \sim \text{Normal}(m = 0, \sigma = 1.5),$$
  $$b_{site} \sim \text{Normal}(m = 0, \sigma = 0.5),$$
  $$c_{arm} \sim \text{Normal}(m = 0, \sigma = 0.5),$$

  with mean m and standard deviation σ.
  Probabilities for superiority of the PCT arm vs. the control arm (shortABx, prescribedABx) will be determined from the posterior sampling distributions of model parameters c_{PCT} and $c_{control}$: Samples will be drawn for c_{PCT} and $c_{control}$ from their posterior distributions, the proportion of sample pairs fulfilling the criterion $c_{PCT} > c_{control}$ (for higher probability of short antibiotic treatment in the PCT-arm vs. the control-arm) or $c_{PCT} < c_{control}$ (for lower probability of antibiotics prescribed at discharge in the PCT-arm vs. the control-arm) will be determined, and it will be assessed if the obtained proportion is above $\gamma = 0.985$ fulfilling the endpoint-specific study success criterion.
  Probabilities for non-inferiority of the PCT arm vs. the control arm (CAE30) will be determined similarly but will additionally include patient bootstrap sampling from the study population and evaluation of the CAE30 success criterion on the level of probabilities for applying the non-inferiority margin M = 10%: (a) draw patient site "i" from study population with replacement, (b) draw samples for model parameters b_i, c_{PCT} and $c_{control}$ from posterior probability distributions, (c)

compute corresponding probabilities $p_{i,PCT}$ and $p_{i,control}$ for CAE30, (d) determine the frequency with which differences between study-arm-specific probabilities fulfill the non-inferiority criterion $p_{i,PCT} < p_{i,control} + M$ (by repeating steps (a) to (c)), and (e) assess if this frequency is above $\gamma = 0.985$ to fulfill the study success criterion.

- Safety analysis (as described in Section 4.4) will also be conducted for the alternative safety endpoint definition CAE30' only counting safety adverse events that occur at least 24 hours after patient randomization. The results will be compared and discussed with the results for counting all safety adverse events after patient randomization.
- For endpoints shortABx, CAE30, and prescribedABx the study-arm specific proportions will be compared and discussed with the corresponding proportions that were obtained for the two run-in studies that were performed for accurate calculation of sample size and required power; these first two ProSAVE pilot-level study batches were named "ProSAVE-1 & -2" (comprising 95 (ITT) and 60 (PP) non-COVID patients not transferred to ICU within 24 hours, patients enrolled from 2020-05-28 to 2021-02-24).
- The endpoints shortABx, CAE30, and prescribedABx will also be analyzed for the sub-population of patients from the pilot study ProSAVE-1 & -2 hospitals, i.e. patients enrolled at Massachusetts General Hospital, North Shore Medical Center or Martha's Vineyard Hospital. The results will be compared specifically with the results previously obtained in the pilot study.
- Estimates and 95% CIs of the endpoints shortABx, CAE30, and prescribedABx will be calculated and visualized site-specifically. Similarities and differences between sites will be discussed.
- The composite safety endpoint CAE30 will be analyzed and discussed regarding its individual constituent adverse event variables.

4.8 Further Secondary and Exploratory Analysis

Secondary and exploratory endpoints will be analyzed according to variable type:

Numerical endpoints:
Distributions of numerical variables will be visualized for all patients and separately for PCT and Control arm. Study arm-specific distributions will be compared with each other by statistical tests (e.g. Wilcoxon rank-sum tests), and by comparing descriptive summary measures (e.g. group medians and means).

Binary endpoints:
2 x 2 contingency tables will be reported for counts stratified according to study arm (PCT, control) and endpoint level (true/false). Associations between study arms and endpoints will be analyzed by statistical tests (e.g. chi-square test or Fisher's exact test). Proportions and two-sided 95% confidence intervals or Bayesian credible intervals will be computed for all patients and separately for PCT and Control arm.

4.9 Data Processing and Quality Assurance

Study data will be collected in an electronic CRF system which is set up and maintained by the study team of the principal investigator at Massachusetts General Hospital (MGH). Collected data will be verified with source data. Data checks will be conducted for consistency (e.g. value ranges, comparisons between different eCRF fields) and completeness.

Prior to final statistical analysis conventional data cleaning and checking procedures will be applied including the assessment of missing data and outliers. Plausibility and consistency checks will be conducted by statistical analyses reviewing the data according to medical background and meaning, e.g. by descriptive statistics such as uni- and bivariate plausibility checks. Non-plausible data entries will be requested for clarification and – if erroneous – for correction.

4.10 Handling of Missing Data

Missing data will be documented and reported in total and per study arm. Missing values will be handled by Bayesian imputation which can be considered as a type of multiple imputation and as the method of choice. In addition, if the amount of missing values turns out to be substantial, e.g. larger than 5%, then the possible impact of missing values on study results will be assessed by sensitivity analysis comparing the obtained study results with results of complete case analysis and with results after worst-case and best-case imputations.

4.11 Extensions and Changes versus Protocol and Previous SAP Versions

Extensions and changes between SAP versions are described in Section ii. “Version History”.

For SAP Version 1.0 issued 2021-12-13 we made the following extensions and changes vs. study protocol version 9.0 [ProSAVE9.0]:

- Revision of the study success threshold γ: $\gamma = 0.9785$.
  The success threshold $\gamma = 0.9785$ was determined in trial simulations together with the specification of thresholds for interim analysis. The set of thresholds was chosen for one-sided statistical significance at the level of 2.5% which corresponds to two-sided testing with 95% confidence level.
- Finalization of the specification of decision thresholds for interim analyses:
  - $\Theta^{F}_{200} = \Theta^{F}_{350} = \Theta^{F}_{500} = \Theta^{F}_{600} = 0.1$,
  - $\Theta^{S,\,CAE30}_{500} = \Theta^{S,CAE30}_{600} = \Theta^{S,\,shortABx}_{500} = \Theta^{shortABx}_{600} = \Theta^{S,\,prescribedABx}_{500} = \Theta^{S,\,prescribedABx}_{600} = 0.9$.
- Explicit specification of hierarchical statistical testing of study endpoints (Section 2.4)
- Characterization of the corresponding study design (Section 2.3.2)
- Detailed specification of interim analysis (Section 2.5)
- Re-ranking of statistical methods for handling missing values (Section 4.10)
- Additional information about randomization with varying block lengths (Section 2.2)
- Re-numbering of secondary endpoints from the third secondary endpoint onwards (Section 4.1)
- Specification of software usage (Section 4.11)

4.12 Data Integrity and Software Validity

All statistical analysis will be conducted and documented by statistical scripts using R, a language and environment for statistical computing and graphics [R Core Team 2018]. Final analysis will be conducted with R 3.5.1 or newer and R-packages of date 2019-12-01 or newer controlled by checkpoint [checkpoint2018]. Scripts will be written by experienced biostatisticians and validated according to the four-eyes-principle by involving a second biostatistician in script reviewing. Final analyses will be conducted in batch mode to assure and document overall consistency of data, conducted analyses and results.

5 References

Study documents

[ProSAVE16.0] PROcalcitonin impact on antibiotic reduction, adverSe events and AVoidable healthcarE costs (ProSAVE): a RCT. Study Protocol Version 16.0, issued on Dec 16, 2022

[GMA Study Folder] https://thermofisher.sharepoint.com/:f:/r/sites/MedicalAffairs-GMAandUS/Shared%20Documents/Clinical%20Studies/ProSAVE?csf=1&web=1&e=JmT5T3

[Biostatistics Study Folder] I:\Bereichsaustausch\Biostatistics\Projects\PCT\PROSAVE_2018

Standard Operating Procedures (see MasterControl for specific referenced versions)

HN-SOP-0001 Quality Manual - Description of the Quality Management System

HN-SOP-0018 Application of Statistical Methods to Generate Data-Based Evidence

HN-SOP-0019 Clinical Trial Control

Standards (see HN-EXT-0166 for specific referenced versions)

ISO 3534-1 Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability

ISO 3534-2 Statistics – Vocabulary and symbols – Part 2: Applied statistics

ISO 3534-3 Statistics – Vocabulary and symbols – Part 3: Design of experiments

Guidelines (see HN-EXT-0166 for specific referenced versions)

ICH-GCP E9 Statistical principles for clinical trials

ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

FDA Guidance for Industry and FDA Staff – Adaptive Designs for Medical Device Clinical Studies

FDA Guidance for Industry and FDA Staff – Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Statistical methods

[Harrel 2001] Harrel, FE: Regression Modelling Strategies. Springer, 2001

[Cowles2013] MK Cowles: Applied Bayesian Statistics. Springer, 2013

[Rajat2021] Rajat et al.: A Sequential Predictive Power Design for a COVID Vaccine Trial, Statistics in Biopharmaceutical Research, 2021 (accepted manuscript)

Clinical references

[deJong2016] de Jong, E., et al., Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. Lancet Infect Dis, 2016. 16(7): p. 819-827

[Huang2018] Huang, D.T., et al., Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection. N Engl J Med, 2018

[Huang2017] Huang, D.T. et al: Design and rationale of the Procalcitonin Antibiotic Consensus Trial (ProACT), a multicenter randomized trial of procalcitonin antibiotic guidance in lower respiratory tract infection. BMC Emergency Medicine (2017) 17:25 1-10

Statistics software

[R Core Team 2018] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[R 3.5.1] R version 3.5.1 (2018-07-02) -- "Feather Spray", Copyright (C) 2018
The R Foundation for Statistical Computing;
with R-packages of date 2019-12-01 controlled by checkpoint

[checkpoint2018] Microsoft Corporation: “checkpoint: Install Packages from Snapshots on the Checkpoint Server for Reproducibility”. R package version 0.4.5, 2018 (https://CRAN.R-project.org/package=checkpoint)