PRODIGY

PRediction of Opioid-induced respiratory Depression In patients monitored by capnoGraphY

Statistical Analysis Plan

Medtronic

	Medtronic
Access	For Internal and External Use
Document Type	Statistical Analysis Plan
Document Title	Prodigy SAP
Clinical Investigation Plan Title	Prodigy: PRediction of Opioid-induced respiratory Depression In patients monitored by capnoGraphY
Clinical Investigation Plan Identifier	COVMOPO0560
Global Sponsor	Covidien LP, a Medtronic company located at: 6135 Gunbarrel Avenue, Boulder, Colorado, 80301 USA
Version Number	Version 2.0 – 18 SEP 2017
Document Reference Number	Clinical Investigation Plan Version 3.0, 29 Nov 2016

Table of Contents

1 Version History

SAP Change History				
Version	**Version Date**	**Description of Change**	**Paragraphs involved**	**Modified by**
1.0	19 Jan 17	First Release	All	Paola Di Stefano
2.0	25 Jan 18	Change in Population set to be analyzed Add new sections 9.3 and 9.3	7.1.3 9	Paola Di Stefano

2 List of Abbreviations and Definitions of Terms

Abbreviation	Definition
ASA PS	American Society Anesthesiologists Physical Status
ASA	Apnea-Sat AlertTM Algorithm
CHF	Chronic Heart Failure
CIP	Clinical Investigation Plan
COPD	Chronic obstructive pulmonary disease
EAC	Episode Adjudication Committee
EC	Ethics Committee
ER	Emergency room
$EtCO_2$	End Tidal CO2
FAS	Full Analysis Set
GEE	Generalized Estimating Equation
ICH	International Conference on Harmonization
ICH E3	ICH guideline E3: Structure and content of clinical study reports
ICH E6	ICH guideline E6: Guideline for Good Clinical Practice
ICH E9	ICH guideline E9: Statistical principles for clinical trials
ICU	Intensive Care Unit
IPI	Integrated Pulmonary Index
OR	Operative Room
ORADE	Opioid Related Adverse Drug Event
PACU	Post Anesthetic Care Unit
PCA	Patient-Controlled Analgesia
PPV	Positive Predictive Value
PPS	Per Protocol Set
PRD	Potential Respiratory Depression event
RD	Respiratory Depression
RR	Respiratory Rate
SAP	Statistical Analysis Plan
SpO_2	Pulse Oximetry
STOP-BANG	**S**noring loudly, **T**iredness in daytime, **O**bserved apnea during sleep, high blood **P**ressure, **B**ody mass index >35 kg/

	m2, **A**ge >50 years, **N**eck circumference >40 cm, and **G**ender

3 Introduction

Opioid therapy is the gold standard for treatment of post-surgical pain in hospital ward but also the majority of non-surgical patients admitted in hospital are exposed to opioids. One of the major opioid side effects includes respiratory depression (RD), which causes alveolar hypoventilation and hypoxemia. Detection of a patient's Respiratory Compromise status before progression can help avert unwarranted outcomes and the possible need for critical care. Typically, only some high-risk patients are monitored by capnography, which assesses real-time ventilation by continuous measuring of SpO2, RR and the concentration of exhaled end tidal carbon dioxide (etCO2). Main risk factors for developing RD have been widely studied in literature: sleep apnea, obesity, snoring, old age, post-surgery, increased opioid dose requirement, concomitant use of other sedating medications, comorbidities like preexisting pulmonary or cardiac disease, PCA use and smoking. The need of a simple tool to stratify patients at risk to develop RD has been underlined by several authors. Such a tool will also help health care providers in selecting the best candidate for capnography monitoring.

This SAP is based on Protocol 3.0, 29 Nov 2016 titled, "Prodigy: PRediction of Opioid-induced respiratory Depression In patients monitored by capnoGraphY". The SAP has been prepared in agreement with Medtronic internal procedures and using the STROBE Statement[1] and International Conference on Harmonization (ICH) guidelines E3, E6 and E9 as guidelines.

4 Study objectives

4.1 Primary objective

The primary objective of this study is to derive and validate a risk assessment tool to identify subjects at risk of having RD while undergoing opioid therapy on the hospital ward.

4.2 Secondary objectives

4.2.1 Secondary objective 1

To compare subjects that will develop RD versus patients that will not develop RD.

4.2.2 Secondary objective 2

To characterize the predictive values of $etCO_2$, RR, SpO_2 and the IPI in predicting RD and ORADE.

4.2.3 Secondary objective 3

To measure health care utilization costs during the study period.

5 Investigation Plan

5.1 Study design

PRODIGY is a prospective, multi-center, post-market interventional, international cohort study. The study population consists of subjects of adult age (≥18 in US and Europe, ≥20 in Japan, ≥21 in Singapore) receiving parenteral opioid therapy for pain while on the hospital ward.

A derivation cohort will be used to derive the risk assessment tool. An internal validation cohort will be used for evaluating the prognostic value of the score for the prediction of RD. Capnography and pulse oximetry monitoring device data will be collected as well as interventions related to respiratory depression. Subjects will be monitored per standard of care.

5.2 Inclusion/Exclusion criteria

All inclusion and exclusion criteria stated in section 8.3 and 8.4 from CIP, must be met for subjects to be eligible for inclusion in the study.

5.3 Overall study design and plan-description

A Subject Screening Log will be filed in the Investigator Site File to document the reason why patients could not be enrolled in the study.

Clinical data will be collected at Enrollment, Capnographic Monitoring Period, 1-month follow-up Visit and Study Exit. Additional data will be collected for Adverse Events, Device Deficiencies and Protocol Deviations.

Data collection requirements are summarized in the following table:

Data	Screening Evaluation	Enrollment Visit	Capnography Monitoring	1 Month Follow up	Study Exit
Informed Consent	X				
Inclusion/Exclusion Criteria Evaluation	X				
Medical History		X			
Demographic & Physical Exam		X			
Vital Signs		X	X		
Supplemental Oxygen Use		X	X		
Surgery Information		X			
Care Pathway			X		
Monitoring Duration			X		
Medications		X	X		
Adverse Events		X	X	X	X
Device Memory Data			X		
Product Information			X		
Device Deficiency			X		
Protocol Deviations	X	X	X	X	X
Healthcare Resource Utilization				X	X

Additional continuous data from the Capnostream monitor will be collected for a maximum of 48 hours to identify potential indicators of respiratory depression.

Monitoring of capnography and pulse oximeter data with the Capnostream monitor will start after opioid therapy has been initiated for subjects that will receive their first opioid therapy dose while on the hospital ward. The monitoring period will start for subjects once they arrive on the ward, for those subjects where opioid therapy was initiated prior to arrival on the hospital ward.

Monitoring may be discontinued after a minimum of 4 hours from the last dose of opioid therapy received or if the subject is discharged from the hospital ward.

6 Determination of Sample Size

The size of the study cohort has been calculated to provide independent samples for derivation and validation cohort. At the study closure, patients will be randomly assigned (2:1) into two groups to create a derivation cohort with 2/3 of the patients and an internal validation cohort with the other 1/3 allowing to the following calculation. According to a generally accepted rule of thumb, at least 10 events per variable are expected to be entered into the logistic regression model. The three following endpoints from RD definition have been taken into consideration to calculate the sample size:

A. RR ≤ 8 bpm for ≥ 3 minutes
B. SpO2 ≤ 85% for ≥ 3 minutes
C. EtCO2 ≥ 60 mmHg for ≥ 3 minutes

From literature the incidence of the above events is at minimum: A = 1.4%, B = 10% and C = 1%. The probability that at least one event among A or B or C occur is the probability (P) of the union of A and B and C minus the probability of their intersection giving the probability of 12%.

Derivation cohort sample size: the size of the derivation cohort has been calculated to provide at least 10 events per variable that we expect to enter into the logistic regression model. Recording at least 120 RD events would allow around 12 predictor variables (as found from literature review) to be entered into logistic regression. Assuming an incidence of 12% of patients with RD episodes and 12-variables for prediction rule, a sample size of 1000 patients is needed for the derivation cohort.

Validation cohort sample size: since the derivation cohort will be 2/3 of the total sample size, the derivation cohort will be 1/3 (500 patients) of the total sample size.

Considering a 10% of withdrawals/dropouts/screening failure, the total sample size needed is 1650 patients.

7 Statistical Methods

7.1 Study Subjects

7.1.1 Disposition of Subjects

Disposition of subjects will be reported following the STROBE Statement Checklist. Number of individuals at each stage of study (number of total assessed for eligibility, number enrolled, number analyzed and number with 1 month follow-up) will be reported. Reason for not participation at each stage will be reported where known.

Table 1 - Number of Subject Screened and Enrollments by Site and Ward
Figure 1 – Subject Enrollment Accrual by Site and Ward Ove Time

Figure 2 – Flow diagram of Patient Disposition

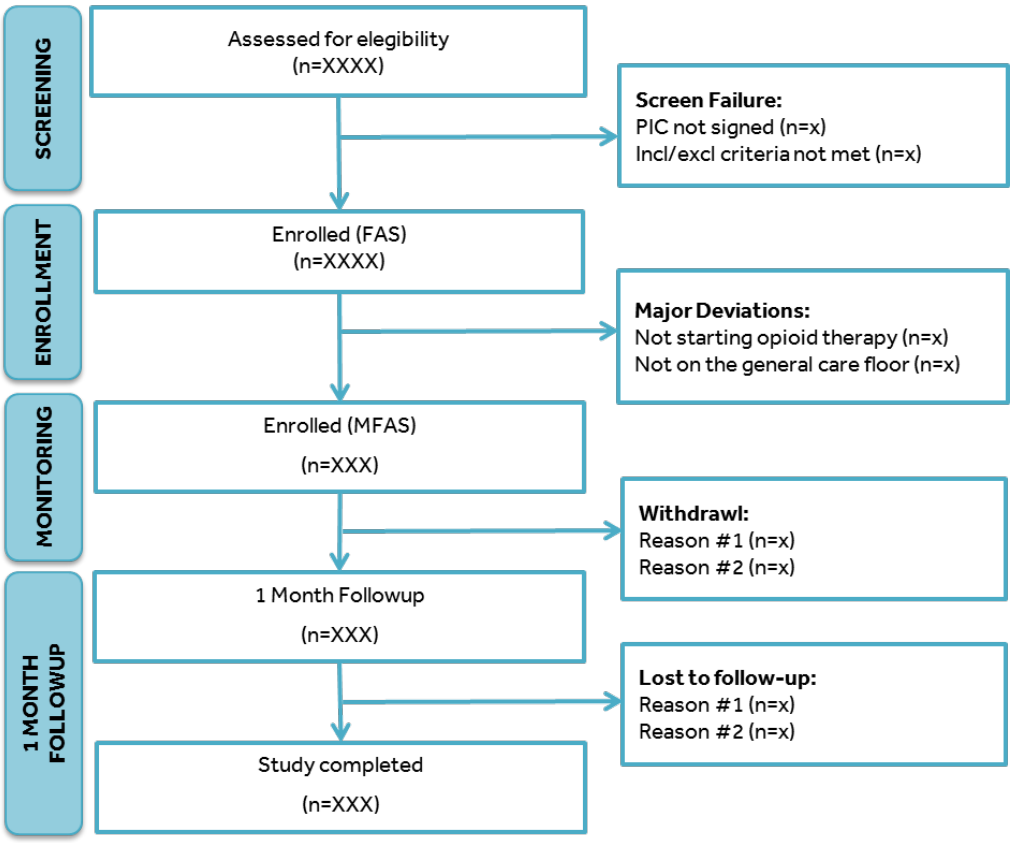

Table 2 - Screening for Eligibility by Inclusion Criteria- FAS
Table 3 - Screening for Eligibility by Exclusion Criteria- FAS

7.1.2 Clinical Investigation Plan (CIP) Deviations

All deviations will be collected in the case report form, with the type of the deviation and the reason for the deviation. All Deviations will be reviewed and classified by the clinical study team. The following tables will describe study deviations:

Table 4 - Protocol Deviation by Reason - FAS
Table 5 - Protocol Deviation by Visit - FAS
Listing 1a. Other deviations - FAS

Protocol deviation will be defined as major protocol deviation if the deviation impacts on the primary objective, such as:

- Enrolled subject did not meet enrolment criteria or eligibility criteria not met (i.e. subject didn't undergo opioid therapy or patient not transferred to the ward).

The following listing will be produced:

Listing1b. Major Protocol deviations - FAS

7.1.3 Analysis Sets

The following subject sets will be used for the analysis:

- The Full Analysis Set (FAS) includes all patients enrolled in the study those sign Patient Informed Consent, fulfill the inclusion and exclusion criteria. The FAS will be used for safety evaluation.

- The Modified Full Analysis Set (MFAS) includes all patients enrolled in the study those sign Patient Informed Consent, fulfill the inclusion and exclusion criteria, start opioid therapy and monitored in the ward. Patients will not be excluded from the MFAS population even if prematurely stops, intermittent use or does not start the monitoring. Patients with major protocol deviations will be excluded from the MFAS. The MFAS will be used to evaluate the primary analysis.

- The Potential Respiratory Depression Population Set (PRDPS) includes all patients in the MFAS and have at least one Episode-File and/or Clinical Event to be adjudicated. The PRDPS will be used for secondary endpoint#2.

The following table shows how each population set will be used for analyses:

Population set	Baseline assessment	Primary and Secondary Endpoints (#1,#3)	Secondary Endpoint#2	Adverse Events
FAS	√			√
MFAS	√	√		√
PRDPS	√		√	

For those patients who have less than the minimum monitoring, or who drop out of the study, the analyses will include all data up to the point of their last data collection.

7.2 General Methodology

At the end of data collection and according to the number of events per variable and number of patients per site, the data-split or the Bootstrap method will be decided. So, all enrolled patients in the MFAS could be randomly assigned to the Derivation Set (DS) and the Validation Set (VS) at a ratio of 2 to 1 or being used all for the derivation and validation. If the method used will be the data-split the randomization will ensure the two cohorts will be not different and they will be compared in terms of baseline characteristics to check possible confounding covariate.

For FAS and PRDPS descriptive statistics will be used to summarize patient characteristics. This will include mean, standard deviation, median, interquartile range, minimum, and maximum for continuous variables, and counts and percentages for categorical variables. It is anticipated that SAS (SAS Institute Inc., Cary, NC, USA) will be used to perform all statistical analyses. Statistical tests will be based on a two-sided significance level of 0.05, and interaction effects evaluated at a significance level of 0.10. Additional exploratory analyses will be conducted as deemed appropriate.

Table 6 - Characteristics of patients in the Derivation and Validation Cohorts– MFAS (if Data-split)
Table 7 - Characteristics of patients – MFAS (if Bootstrap)

7.3 Center Pooling

Enrollment at any single site will be limited to 330 patients (20%) to limit bias. Based on the clinical practice in the opioid used and number of subjects enrolled per center/country, an investigation of center/country effect will be performed including an independent variable for center/country in the model as well as summary statistics if needed.

7.4 Handling of Missing Data and Dropouts

The choice of the imputation method for missing data will depend on the amount and on the pattern of missingness in the data and the type of the imputed variable. The proc MI procedure will be used to output the missing data patterns for the variables. The outcome variable will not be imputed since the CEC will adjudicate all patients with at least one RD.

7.5 Adjustments for Multiple Comparisons

No adjustments for multiple comparisons or multiple look at data will be performed.

7.6 Derived variables

The statistical evaluations will use some derived variables as listed below. All additional derived variables or changes on the variable listed below are reported in the Statistical Programming Requirements document.

New Variable§	**Derivation**
Monitoring Exposure (minutes)	End monitoring date/time – Start monitoring date/time
Study Exposure (days)	(Study Exit date – date of enrollment)/30.44
BMI35	If BMI >35 then BMI35=1 otherwise BMI35=0
Obese	If BMI >30 then Obese=1 otherwise Obese=0
AGE50	If Age >50 then AGE50=1 otherwise AGE50=0
AGE65	If Age >65 then AGE65=1 otherwise AGE65=0
STOP BANG SCORE	The score is the sum of the following points for each patient: If Snoring loudly= Yes then add 1 point If Tiredness in daytime= Yes then add 1 point If Observed apnea during sleep= Yes then add 1 point If Treated for High blood Pressure= Yes then add 1 point If Body mass index >35 (kg/m2) then add 1 point If Age >50 years then add 1 point If Neck circumference >40 cm (16 inches/17inches according to gender) then add If Gender = Male then add 1 point
Opioids Tolerant §	The FDA definition will be used: A patient is considered opioid tolerant if for at least 1 week he/she has been receiving one of the following: • 60 mg oral morphine/day • 25 mcg transdermal fentanyl/hour • 30 mg oral oxycodone/day • 60 mg oral hydrocodone/day • 8 mg oral hydromorphone/day • 25 mg oral oxymorphone/day • Equi-analgesic dose of any other opioid The Morphine Milligrams Equivalent will be used.
STOP BANG SCORE Class	If 0≤STOP-BANG score ≤2 then STOP BANG SCORE Class= 1 labeled as "Low risk of OSA" If 3≤STOP-BANG score ≤4 then STOP BANG SCORE Class= 2 labeled as "Intermediate risk of OSA" If 5≤STOP-BANG score ≤8 then STOP BANG SCORE Class= 3 labeled as "High risk of OSA"
High Risk Surgery	If occurred within 24 hours
Known or suspected sleep-disordered breathing§	if "Yes to OSA (from Medical History) + YES to first 4 questions of the Apnea History
PCA or epidural or intrathecal therapy	PCA with route as epidural or intrathecal from CM
Multiple opioid or concurrent CNS/sedating medication	According to the Medication coding *(benzodiazepines, sleep aids, muscle relaxant, etc.)
Opioid dosage	According to the Medication coding * (>30mg oral morphine per day or equivalent)
Major organ failure	According to the coding °
Diabetes	if "Yes to Diabetes type I or II"

New Variable§	Derivation
Chronic heart failure or other significant cardiac disease	if "Yes to CHF (from Medical History) + YES to any of the followings from Medical History: Coronary artery disease, Myocardial infarction, Stroke, Transient ischemic attack (TIA), Mitral valve disease, Aortic valve disease, Aortic aneurysm, Cerebral aneurysm, Peripheral vascular disease, Hypertension.
Smoke (> 20 packs per year)§	if "Yes to Smoke (any usage)"

* WHO Drug Dictionary, Version 2009MAR
° MedDRA dictionary, version 20.0
§ all details or potential changes in derived variable are listed in the Statistical Programming Requirements Document.

7.7 Demographic and Other Baseline Characteristics

Descriptive statistics will be used to summarize demographic and baseline characteristic variables for FAS, MFAS and PRDPS. This will include mean, standard deviation, median, minimum, and maximum for continuous variables, and counts and percentages for categorical variables. Demographic and Baseline variables will be collected through: Medical history coded using the MedDRA dictionary version 20.0 (including surgical, apnea, substance use, and medical therapy such as oxygen use and opioid use), Demographic information, Vital signs, Supplemental oxygen use, Physical exam, Surgery information (post-surgical subjects only) and Medication. The following tables will be reported according to the mock tables presented in the Statistical Programming Requirements document:

Table 8 - Subject Demographics – FAS, MFAS, PRDPS
Table 9 - Subject Physical and Clinical Assessment – FAS, MFAS, PRDPS
Table 10 - Vital Signs/ Resting Baseline per Standard of Care FAS, MFAS, PRDPS
Table 11 - Medical Therapy/Treatments - FAS, MFAS, PRDPS
Table 12 - Surgical Information (for post-surgical patient only) - FAS, MFAS, PRDPS
Table 13 - Medical History Abnormalities by Primary System Organ Class and Preferred Term - FAS, MFAS, PRDPS
Table 14 - Medical History - FAS, MFAS, PRDPS (for a subset of terms if needed)
Table 15 - Surgical History - FAS, MFAS, PRDPS
Table 16 - Apnea History - FAS, MFAS, PRDPS
Table 17 - Substance Use/Smoking History - FAS, MFAS, PRDPS
Table 18 - Medical Therapy History - FAS, MFAS, PRDPS

Medication will be coded using the WHO Drug Dictionary, Version 2009MAR and with the following summary tables:

Table 19 – Medication by WHO Drug Dictionary - FAS
Table 20 – Medication by MedDRA dictionary - FAS

7.8 Treatment Characteristics

Extent of exposure in the population is characterized according to the duration of opioid therapy with monitoring.

Duration of **Opioid Exposure before Monitoring** will be measured in minutes from the start of the Opioid therapy and the starting of the Monitoring: Duration of Opioid Exposure before Monitoring exposure (minutes) = (Start monitoring date/time – Start Opioid therapy date/time). The extent of Opioid Exposure before Monitoring will be presented in a summary table and supporting data listing.

Table 21 – Opioid Exposure before monitoring - FAS, MFAS, PRDPS

Duration of **Monitoring Exposure** will be measured in minutes from the start of the monitoring through and including the time of monitoring end: Duration of monitoring exposure (minutes) = (End monitoring date/time – Start monitoring date/time). Extent of Monitoring exposure will be presented in a summary table and supporting data listing.

Table 22 – Monitoring Exposure - FAS, MFAS, PRDPS

The monitoring exposure will be also described in terms of the quality check, reporting EtCO2 and SpO2 gap overall and during the first 8 hours.

Duration of **Study Exposure** will be measured in days starting from the point of enrollment (informed consent completed and inclusion/exclusion criteria confirmed per the screening evaluation) through and including the time of study exit: Duration of study exposure (days) = (Study Exit date – date of enrollment). Extent of study exposure will be presented in a summary table and supporting data listing.

Table 23 - Study Exposure - FAS, MFAS, PRDPS

Descriptive statistics will be used to summarize study device information for FAS:

Table 24 - Study Device Information - FAS

7.9 Interim Analyses

Interim analyses are not planned for this study.

7.10 Evaluation of Objectives

In this section a detailed information about each objective is included together with calculations and derivations of outcome parameters (see table in section 7.6), analysis methods, datasets analyzed (FAS, MFAS or PRDPS) and additional analyses where applicable.

7.10.1 Primary endpoint

The primary endpoint that will be used to derive and validate the risk score assessment is the RD episode, defined as a clinical diagnosis made after reviewing monitoring data in conjunction with the clinical data and consistent with accepted pathophysiologic mechanisms. The primary endpoint used to derive the score will be the occurrence of RD episodes captured by continuous capnography and pulse oximetry measurements recorded on the Capnostream device memory data in conjunction with the clinical data as reported by the Investigators. RD determination will be validated by an independent Clinical Event Committee (CEC). The CEC will use both clinical data and monitor parameters specified below as a guideline; if determination of an RD episode is clinically appropriate but outside of the guidelines a rationale will be provided. Data suggestive of an RD episode include any of the following:

RR ≤ 5 breaths for ≥ 3 minutes
SpO2 ≤ 85% for ≥ 3 minutes

etCO2 ≤ 15 or ≥ 60 mmHg for ≥ 3 minutes
Apnea episode lasting > 30 seconds
Any respiratory Opioid-Related Adverse Event (rORADE)

The assessment of RD episodes will be made by the analysis of the Capnostream and clinical data for each patient. For each patient all potential RD episodes will be derived using the Detection Tool (see section 9.3) and will be submitted to the CEC according to a Priority List (see section 9.4) in order to adjudicate the potential RD episodes starting from RD episodes most likely to be RD. The CEC adjudication will stop the adjudication at the first RD occurring per patient (all details on CEC process are reported in the CEC Charter version 3.0).

All potential RD episodes in the study will be always rated by 3 coders (see detail in the Clinical Events Committee Charter Template, Version 3.0).

Patients of the MFAS will be randomly assigned (2:1) into respectively the derivation and the validation cohort. In case of low enrollment rate or low episode rate or according the outcome of a preliminary assessment of the events per variable or unbalanced enrollment per site/country the MFAS could not be randomized but used for model building and the bootstrap method, deriving 500 computer-generated samples by random selection with replacement for model validation[3] (see section 9.2 for details).

7.10.1.1 Deriving the model

Note for Programmer: The analysis for the primary objective will be performed on a Dataset with one row per patient with RD episode as validated by CEC as the dependent variable (at least one RD) and baseline predictors as independent variables. Potential predictors will be reported on continuous scale or grouped into categories.

The logistic model will be used for bivariate e multivariate odds ratios (ORs) and 95% confidence intervals (95% CIs) will be estimated for each potential predictor. The following steps will be followed to derive the score for each predictor according to categories[4]:

- Determine categories for each risk factors (the most will be 0/1)
- Determine reference class for each risk factor (for dummy the reference category is 0)
- Calculate mid-values to represent categories
- Determine multiplier β
- Determine point score for each risk factor category

If **Data Split** approach:

- Repeat multivariate model on the validation set
- Calculate performance measures
- The "Optimism" is the difference between the performance measures from validation and derivation set

If **Bootstrapping** approach:

- 500 bootstrapped replications of the multivariate model

- Calculate performance measures bootstrapped
- The difference between the bootstrapped performance and the model performance will be averaged to obtain an estimation of the "optimism" of the model. The optimism will be used to validate the model.

Table 25 - RD Risk Score Derivation identified by logistic – MFAS

Predictors	Bivariate analysis OR (95% CI) (N=XXXX)	Multivariate analysis OR (95% CI) (N=XXXX)	Coefficient	Adjusted Coefficient	Risk score point
Char#1	x.xx (x.xx- x.xx)	x.xx (x.xx- x.xx)	x.xxx	x.xx (x.xx- x.xx)	xx
Char#2	x.xx (x.xx- x.xx)	x.xx (x.xx- x.xx)	x.xxx	x.xx (x.xx- x.xx)	xx
…					
Char#N	x.xx (x.xx- x.xx)	x.xx (x.xx- x.xx)	x.xxx	x.xx (x.xx- x.xx)	xx

The collinearity between categorical variables will be also tested by means the Cramer V test for nominal variables and Kendal's tau-b for ordinal variables. The multivariate logistic regression model will be performed using a stepwise selection procedure in which the presence of event will be the dependent variable. Independent predictors will be entered in to the model if a significant association, defined as p≤0.05, will be identified from bivariate analysis and, to avoid over-fitted and unstable model the correlation coefficient between them should be less than 0.25.

The predictive risk score for RD, based on the best model selected from analysis above, will be calculated by multiplying each β coefficient taken from the final multivariate model by 10 and rounding to the nearest integer. The integers will then be added together to produce an overall RD risk score for each patient. The resulting continuous distribution of total risk scores across all patients in the derivation set will be then stratified into categories of points (depending on the total sample the quintiles or tertiles or optimal cutoff points) that grouped patients according to the level of risk.

7.10.1.2 Validating model on validation cohort

Model prediction performance will be assessed as for the derivation cohort using calibration (the agreement between predicted and observed outcome) and discrimination (the ability to separate patients with and without the outcome) and the R^2 will be reported as an overall measure for discrimination and calibration. The Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and the Receiver Operating Characteristic (ROC) curves will be used to visualize the overall accuracy of the model (see section 9.1 for detail on aspect to be checked for model validity).

Table 26 - Sensitivity, specificity, PPV and NPV of the Risk Score to predict high risk subjects – Validation Cohort - FAS

Accuracy Measure	Optimal Cut-off %(IC)
Sensitivity	X% (Y%-Y%)
Specificity	X% (Y%-Y%)
PPV	X% (Y%-Y%)
NPV	X% (Y%-Y%)
…	…

Figure 3 – Receiver Operating Characteristics curve on Validation Cohort-FAS
Figure 4 – Agreement between observed frequency and predicted probability on Validation Cohort- FAS

The discriminatory performance of the model will be validated by comparing the receiver-operating characteristic (ROC) curve analysis in the derivation set with that in the validation set.

Figure 5 – Comparison between ROC curves on Derivation and Validation Cohorts-FAS
Figure 6 – RD rate according to the Risk Score in Derivation and Validation Cohorts-FAS

7.10.2 Secondary endpoints

For secondary endpoints the MFAS population will be used for analysis.

7.10.2.1 Secondary endpoint#1

The secondary objective#1 will be evaluated comparing subjects that will develop RD versus patients that will not develop RD in terms of:

- Incidence of AE and actions taken.
- Healthcare resource utilization (including hospital length of stay, 30 days readmission rate and primary diagnosis upon readmission).
- Subject mortality at 30 days.

This secondary objective will be explained in the Health Economic Analysis Plan (HEAP).

7.10.2.2 Secondary endpoint#2

The secondary objective#2 will characterize the predictive values of etCO2, RR, SpO2 and the IPI[5] in predicting rORADE. To evaluate the predictive values of the Capnostream monitoring parameters for rORADE, the Sensitivity, Specificity, PPV and NPV will be calculated. rORADE events will be considered the Gold standard and the CS20p monitoring parameters as the measurement to be tested. The population used for this endpoint is PRDPS adjudicated as RD. All Clinical Events occurring during not monitored period will be reported but excluded from the table below. For each RD adjudicated by the CEC, the Capnostream monitoring will be checked if any of the parameter meets at least once the thresholds (RR ≤ 6 breaths for ≥ 3 minutes, SpO2 ≤ 85% for ≥ 3 minutes, EtCO2 ≥ 60 mmHg for ≥ 3 minutes or IPI < 3,4,5,6). The table below shows the distribution of the true and false positive and negative for each monitored parameter:

Test		Gold Standard		
Monitoring Parameter		**ORADEs/RDs**		
RR		**YES**	**NO**	**Tot**
≤ 5 breaths for ≥ 3 minutes	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
Monitoring Parameter		**ORADEs/RDs**		
SpO2		**YES**	**NO**	**Tot**
≤ 85% for ≥ 3 minutes	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
Monitoring Parameter		**ORADEs/RDs**		
etCO2		**YES**	**NO**	**Tot**
≤ 15 or ≥ 60 mmHg for ≥ 3 minutes	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
Monitoring Parameter		**ORADEs/RDs**		
IPI		**YES**	**NO**	**Tot**
<3	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
IPI		**YES**	**NO**	**Tot**
<4	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
IPI		**YES**	**NO**	**Tot**
<5	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
IPI		**YES**	**NO**	**Tot**
<6	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N
Monitoring Parameter		**ORADEs/RDs**		
Any Monitoring Parameter		**YES**	**NO**	**Tot**
RR ≤ 6 breaths for ≥ 3 min or SpO2 ≤ 85% for ≥ 3 min or EtCO2 ≥ 60 mmHg for ≥ 3 min or IPI < 3,4,5,6	**YES**	TP	FP	TP+FP
	NO	FN	TN	FN+TN
	Tot	TP+FN	FP+TN	N

Where:

- N will be all RDs adjudicated by the CEC.
- TP= true positive are all rORADEs and parameter over/below the threshold.
- FN= false negative are all rORADEs but no parameter over/below the threshold.
- FP= false positive are all no rORADE but parameter over/below the threshold.
- TN=true negative are all no ORADEs and no parameter over/below the threshold.

Measure of the accuracy will be:

- Sensitivity: {TP/(TP+FN)},
- Specificity: {TN/(FP+TN)},
- Positive predictive value: {TP/(TP+FP)},
- Negative predictive value: {TN/(TN+FN)}

Table 27 – Sensitivity, specificity, PPV and NPV of RD and ORADE – PRDPS

Accuracy Measure	RR %(IC)	SpO2 %(IC)	etCO2 %(IC)	IPI(<3) %(IC)	IPI(<4) %(IC)	IPI(<5) %(IC)	IPI(<6) %(IC)
Sensitivity	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)
Specificity	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)
PPV	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)
NPV	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)	X% (Y%-Y%)
…	…	…	…	…	…	…	…

Figure 7 – Receiver Operating Characteristics curve for RR- PRDPS
Figure 8 – Receiver Operating Characteristics curve for SpO2- PRDPS
Figure 9 – Receiver Operating Characteristics curve for EtCO2- PRDPS
Figure 10 – Receiver Operating Characteristics curve for IPI<3- PRDPS
Figure 11 – Receiver Operating Characteristics curve for IPI<4- PRDPS
Figure 12 – Receiver Operating Characteristics curve for IPI<5- PRDPS
Figure 13 – Receiver Operating Characteristics curve for IPI<6- PRDPS

7.10.2.3Secondary endpoint#3

The secondary objective#3 is to measure health care utilization costs during the study period. The Health Economic Analysis Plan (HEAP) will be developed for this secondary endpoint and other additional analyses.

7.11 Safety Evaluation

Adverse events will be presented using the MedDRA coding and with the following summary tables and supporting data listing:

Table 28 – Adverse Event by Primary System Organ Class and Preferred Term – FAS
Table 29 – Adverse Event Seriousness and Relatedness– FAS
Table 30 – Individual Adverse Events for which Sponsor and Investigator disagreed – FAS
Listing2- Individual Adverse Events - FAS
Table 31 – Death Summary – FAS (if appropriate)
Listing 3- Death – FAS (if appropriate)

7.12 Changes to Planned Analysis

The analysis described in the CIP could differ from that presented in this SAP due to data availability. In case a lower rate of enrollment will occur or any other unexpected

issue which reduce the sample size or event rate a new strategy will be adopted to protect the derivation cohort and to allow at least the derivation of the score[6,7] (see section 9.2 for details).

8 Validation Requirements

All collected data will be reviewed for completeness, correctness and consistency. In case of issues, queries will be sent to the investigator to complete, correct or comment the data. To ensure the quality of the results provided for the study in the form of tables, listings and figures, and the derived datasets the following processes are used:

- Statistical programming and analysis will be done by qualified programmer(s) and statistician(s) following applicable procedures and best practices.
- The derived datasets will be validated by a second programmer or statistician.
- The tables will be validated by a second programmer or statistician.
- Statistical results will be reviewed and confirmed by a second statistician.

The entire set of tables, listings, and figures (TLF) will be 100% checked for accuracy, completeness, and consistency prior to inclusion in the interim or final clinical study report. According to Medtronic SOPs the level II validation (the peer reviewer reviews the code; where appropriate, performs manual calculations or simple programming checks to verify the output) will be implemented for both Datasets and TLFs.

9 References

1. http://www.strobe-statement.org/index.php?id=available-checklists
2. Mary L. McHugh: Inter rater reliability: the kappa statistic (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/?report=reader)
3. Frank e. Harrell, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy and measuring and reducing errors, Statistics in Medicine, vol. 15,361-387 (1996)
4. Presentation of multivariate data for clinical use: The Framingham Study risk score functions, Lisa M. Sullivan, Joseph M. Massaro, Ralph B. D'Agostino Sr., Tutorial in Biostatistics, Stat. Med 2004
5. Ronen M., Smart respiratory monitoring: clinical development and validation of the IPI™ (Integrated Pulmonary Index) algorithm, Journal of Clinical Monitoring and Computing, pp 1-8 mar 2016
6. Menelaos P., How to develop a more accurate risk prediction model when there are few events, BMJ 2015
7. Royston P., External validation of a Cox prognostic model: principles and methods, BMC Medical Research Methodology 2013
8. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin. 1971; 76(5):365–377.

9.1 Measures for model validity

The measures calculated will be:

- The **concordance**: the *c* statistic. For binary outcomes, *c* is identical to the area under the receiver operating characteristic (ROC) curve; *c* varies between 0.5 and 1.0 for sensible models (the higher the better).

- The **calibration slope** is the regression coefficient b in a logistic model with the predictive score as the only covariate: logit(mortality) = a+ b * predictive score. Well-calibrated models have a slope of 1, while models providing too extreme predictions have a slope less than 1.

- The **Brier score** (or average prediction error) is calculated as Sum(y_i - p_i)^2/n, where y_i denotes the observed outcome and p_i the prediction for subject *i* in the data set of *n* subjects.

- The ***D* statistic** is a scaled version of the model chi-square, which is a function of log-likelihood

- The **R^2** as a measure of explained variation.

The table below summarize the three aspects to be checked for model validity:

Aspect	Measure	Characteristics
Calibration	Calibration plot	Visual impression of observed frequencies vs. predicted probabilities
	Slope	Estimate of extremeness of predicted probabilities
	Intercept	Estimate of systematically too high/low predicted probabilities
	E_{avg}	Average absolute difference between observed frequencies and predicted probabilities
	Hosmer-Lemeshow Statistic	Test for 'goodness-of-fit', i.e. deviance of grouped observed outcomes and predicted outcomes
Discrimination	Boxplot of predicted probabilities	Visual impression of spread in predicted probabilities; relies on adequate calibration
	c-statistic (ROC area)	Summary of quality over a range of threshold values
Clinical usefulness (threshold value required)	Accuracy	Percentage of patients correctly classified, given a certain threshold value
	Sensitivity	Percentage of patients with the outcome correctly classified as diseased
	Specificity	Percentage of patients without the outcome correctly classified as non-diseased
	Decrease in weighed false classifications	Model and reference policy are compared by weighing patients falsely classified as diseased and non-diseased according to relative severity

9.2 Bootstrap Method

Bootstrapping will be considered as an alternative method of internal validation that involves taking a large number of samples with replacement from the original sample. Bootstrapping provides unbiased estimates of predictive accuracy and it has the advantage to use the entire dataset for model development. A total of 500 bootstrapped replications will be performed. For each replication the univariate and multivariate models will be run. According to results, the R^2 as the average of R^2 from each replication will be calculated. The difference between average bootstrapped R^2 and model R^2 will represent the "optimism" of the model compared to the bootstrapped models. The optimism will consider as the validation of the model.

9.3 Detection Tool

The Detection Tool analyzes each patient's monitoring using an algorithm based on the device episodes thresholds as stated in the primary endpoint. The Detection Tool provides for each patient multiple Episodes. Since there could be multiple episodes in a short time period around each Episode, the Detection Tool is set to aggregate episodes with at most 30 minutes time difference between them in one file. In addition the episodes file consists of time window of 30 min before and after the first and last episode.

9.4 Priority List

The CEC will stop the adjudications at the first RD for each patient, while they will adjudicate all Clinical Events. In order to optimize CEC time the following priority list will be used to prioritize some Episode-File to be review first:

Priority 1: all clinical events with/without any episode from device data. In case multiple Clinical events per patient the CEC will review them using the time of occurrence, from first clinical event to the last.

Priority 2: Multiple episodes involving at least two parameters among SpO2, EtCO2, RR and Apnea. In case of multiple Episode-Files with this priority the CEC will review first the Episode-Files with the higher number of Episodes within the Episode-Files.

Priority 3: Multiple Episodes of the same parameter. In case of multiple Episode-Files with this priority the CEC will review using this order: SpO2, EtCO2, RR and Apnea

Priority 4: Any overnight Episodes of the same parameter. In case of multiple Episode-Files with this priority the CEC will review using this order: SpO2, EtCO2, RR and Apnea