STATISTICAL ANALYSIS PLAN

Perioperative Predictive Value of Physical Activity on Short-, Mid- and Long-term Mortality: A Machine Learning Approach

ClinicalTrials.gov Identifier	NCT07603622
SAP Version	1.0
SAP Date	26 08 2026
Study Title	Predictive value of preoperative functional capacity In Mortality Estimation: assessed using a machine learning approach
Principal Investigator(s)	**Max Bell, MD, Associate Professor** Perioperative Medicine and Intensive Care, Karolinska University Hospital Department of Physiology and Pharmacology, Karolinska Institute
SAP Authors	**Arman Valadkhani, MD, PhD** Perioperative Medicine and Intensive Care, Karolinska University Hospital Department of Physiology and Pharmacology, Karolinska Institute **Martin Jacobsson, MD, PhD** Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology. **Max Bell, MD, Associate Professor** Perioperative Medicine and Intensive Care, Karolinska University Hospital Department of Physiology and Pharmacology, Karolinska Institute

This Statistical Analysis Plan was finalised and locked prior to data analysis. Any subsequent amendments will be documented in the amendment history.

Table of Contents

Table of Contents

Table of Contents *2*

1. Administrative Information *3*

2. Introduction and Background *4*

2.1 Clinical Background **4**

2.2 Objective **4**

3. Study Methods *5*

3.1 Study Design **5**

3.2 Study Population and Recruitment **5**

3.3 Inclusion and Exclusion Criteria **5**

3.3.1 Inclusion Criteria 5

3.3.2 Exclusion Criteria 5

3.4 Follow-up and Data Collection **5**

4. Measures and Variables *6*

4.1 Predictor of Interest **6**

4.2 Outcome Definitions **6**

4.3 Predictors **6**

5. Statistical analysis *8*

5.1 Sample size for prediction model development **8**

5.2 Primary analysis **8**

5.2.1 Baseline Characteristics 8

5.2.2 Model development 8

5.4 Sensitivity analysis **9**

5.5 P-values **9**

5.6 Missing Data **9**

6. Preregistration and Study Reporting *10*

6.1 Preregistration **10**

6.2 Study Reporting **10**

References *11*

1. Administrative Information

Item	Detail
Full Study Title	Predictive value of preoperative functional capacity In Mortality Estimation: assessed using a machine learning approach
Short Title / Acronym	**Periop**erative **Pre**dictive Value of **P**hysical **A**ctivity on Mortality (**PeriopPrePA**)
ClinicalTrials.gov ID	NCT07603622
Study Design	Prospective observational cohort study
Number of Centers	Two (dual-center)
Study Population	Adults (≥18 years) undergoing elective non-cardiac surgery
Total Target Sample Size	Approximately 50,000 participants
Predictor of interest	Metabolic equivalents of task (MET)
Primary Outcomes	30-day all-cause mortality; 90-day all-cause mortality; 365-day all-cause mortality
SAP Authors	**Arman Valadkhani, MD, PhD** Perioperative Medicine and Intensive Care, Karolinska University Hospital Department of Physiology and Pharmacology, Karolinska Institute **Martin Jacobsson, MD, Associate Professor** Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology. **Max Bell, MD, Associate Professor** Perioperative Medicine and Intensive Care, Karolinska University Hospital Department of Physiology and Pharmacology, Karolinska Institute
Ethics Approval Reference	Swedish Ethical Review Authority, Sweden Dnr 2024-01664-01
Date of Data Lock	Data were extracted and frozen on 02 02 2026 The statistical analysis plan was developed and locked on 26 08 2026, prior to any formal analysis of the study outcomes. No outcome data were examined during SAP development.
SAP Lock Date	26 08 2026

Good Clinical Practice Statement
This SAP has been written and finalized prior to data analysis in accordance with good statistical practice and the TRIPOD + AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence) reporting guidelines. Any post-hoc deviations from this plan will be clearly identified as such in all manuscripts.

2. Introduction and Background

2.1 Clinical Background

For background, see Clinicaltrial.gov application NCT07603622.

2.2 Objective

To develop and internally validate a machine learning model for predicting short-, mid-, and long-term postoperative mortality in adults undergoing elective non-cardiac surgery, and to evaluate whether the addition of preoperative self-reported functional capacity, measured using Metabolic Equivalent of Task (MET) categories, improves predictive performance beyond a baseline model comprising established preoperative predictors.

3. Study Methods

3.1 Study Design

This is a prospective observational cohort study conducted at two academic medical centers.

3.2 Study Population and Recruitment

Data were prospectively collected at two sites within the Karolinska University Hospital network, Solna and Huddinge, Stockholm, Sweden, between 2020 and 2024. Both sites are tertiary academic hospitals with established perioperative research infrastructure, and together they contribute a heterogeneous surgical population, supporting the generalizability of the developed model.

3.3 Inclusion and Exclusion Criteria

3.3.1 Inclusion Criteria

- Adults aged ≥ 18 years at the time of surgery
- Elective (non-emergency) surgical procedure
- Non-cardiac surgery (all surgical specialties except cardiac surgery)

3.3.2 Exclusion Criteria

- Obstetric surgery
- Transplant surgery
- Planned day-case surgery (procedures with intended same-day discharge)
- Anesthetic monitoring episodes not involving a surgical procedure (e.g., anesthetic support for central venous catheter insertion)
- Urological brachytherapy
- Stereotactic radiosurgery (Gamma Knife or equivalent)
- Repeat or subsequent surgical procedures during the study period — only the index (chronologically first eligible) procedure per patient is included in the analysis

3.4 Follow-up and Data Collection

All data collected were registered prospectively (prior to the at-risk period for all study outcomes) in two electronic health record systems (TakeCare and Orbit) and merged into a study database. Comorbidities were recorded as the 10th revision of the International Classification of Diseases (ICD-10) codes. When ICD-codes were not available, they were recorded based on the preoperative assessment made by the attending anesthesiologist.

4. Measures and Variables

4.1 Predictor of Interest

Before index surgery, all patients reported their functional capacity using standardized MET activity thresholds, see Table 1. These reported MET scores could be amended by the attending anesthesiologist during the preoperative assessment if they were deemed to be unrepresentative of the patient's functional capacity. Functional capacity using the MET score was registered as a five-level categorical variable (MET 1, 2 to 3, 4 to 5, 6 to 8, 9 or greater).

Table 1. The level of physical fitness that corresponds to each level of the MET variable

MET	Level of physical fitness
≥9	I can do **more demanding physical activity**, such as normal jogging (7 minutes per km), carrying moving boxes, or similar, or more than this
6-8	I can do **lighter physical activity**, such as slow jogging (9 minutes per km), easy swimming, mowing the lawn with a hand mower, shoveling snow, or similar
4-5	I can do **heavier household work**, such as vacuuming or mopping floors, climbing a flight of stairs (15 steps) **without stopping**, or similar
2-3	I can do **lighter household work**, such as doing the dishes, walking a short distance on level ground at a slow pace, or similar
1	I get out of breath from the slightest effort, for example, walking from the bed to the bathroom

Metabolic equivalent of task (MET) as a five-level categorical variable was used to assess preoperative physical fitness.

4.2 Outcome Definitions

Table 2. Outcome Definitions

Outcome	Definition and Ascertainment Method
30-day all-cause mortality (Primary)	Death from any cause within 30 days after index surgery. Ascertained via electronic health record database which is linked to national death registry.
90-day all-cause mortality (Primary)	Death from any cause within 90 days after index surgery. Ascertained via electronic health record database which is linked to national death registry.
365-day all-cause mortality (Primary)	Death from any cause within 365 days after index surgery. Ascertained via electronic health record database which is linked to national death registry.

4.3 Predictors

The following predictors will be collected at the time of preoperative assessment.

Demographic predictors

- Age (years)
- Weight (kg)
- Body mass index (BMI) (kg· m⁻2)
- Sex

Comorbidities

- Hypertension
- Ischemic heart disease
- Atrial fibrillation
- Heart failure
- Ischemic stroke

- Transient ischemic attack (TIA)
- Peripheral artery disease
- Chronic obstructive pulmonary disease (COPD)
- Diabetes mellitus type 1
- Diabetes mellitus type 2
- Hyperlipidemia
- Chronic kidney disease

Preoperative Biochemical labs

- Hemoglobin ($g \cdot L^{-1}$)
- Creatinine ($\mu mol \cdot L^{-1}$)
- Albumin ($g \cdot L^{-1}$)

Preoperative risk assessment variables

- American Society of Anesthesiologists (ASA) Physical Status Classification
- Revised Cardiac Risk Index (RCRI)

Perioperative characteristics

- Type of surgery
- Anesthetic technique (general, epidural, spinal, nerve block, local)
- Surgical risk (classification according to the European Society of Cardiology's guidelines for non-cardiac surgery, 2022) [1]
- Duration of surgery, in min
- Duration of anesthesia, in min

5. Statistical analysis

5.1 Sample size for prediction model development

All patients who underwent surgery between 2020 and 2024 were screened for eligibility. This created a fixed study population size and precluded the need for a priori sample size calculation.

5.2 Primary analysis

5.2.1 Baseline Characteristics

Baseline characteristics will be summarized and stratified by MET status, using median and interquartile range (IQR) for continuous variables and frequency with percentage for categorical variables.

5.2.2 Model development

Train and Test split

The complete dataset will be randomly partitioned into a training set (60%) and a held-out test set (40%) using stratified sampling by outcome status to ensure comparable event rates across both partitions. This is because the incidence of the outcome will be rare, between approximately 0-5 %.

Predictor selection

The baseline model will include age, ASA physical status classification, BMI, and surgical risk, selected a priori based on domain knowledge as established perioperative risk factors [2].

Model type selection

To identify the most predictive baseline model for 30-, 90- and 365-day postoperative mortality, three candidate models will be evaluated: a parametric accelerated failure time model (AFT) with a log-logistic distribution, a random survival forest, and a XGBoost model. For each candidate model, a pre-specified hyperparameter search space will be defined prior to any model fitting.

Model selection will be performed using nested k-fold cross-validation on the training set. In the inner loop, hyperparameters will be optimized for each candidate model using 3-fold cross-validation, guided by the concordance index as the selection criterion. In the outer loop, the performance of each optimally tuned model will be estimated using 5-fold cross-validation, yielding a bias-corrected estimate of performance. The inner loop serves only to rank hyperparameter combinations rather than provide unbiased performance estimates. All cross-validation folds will be stratified by outcome status to ensure consistent event rates across partitions, given the low incidence of the outcome.

If the incidence of the outcome is too rare to perform CV for model selection, the same model as the adjoining outcome horizon will be used.

Hyperparameter tuning

The candidate model achieving the highest mean concordance index (C-index) across the outer folds will be selected as the final model type. Following model type selection, a final hyperparameter tuning step will be performed on the complete training set using 5-fold cross-validation, identifying the globally optimal hyperparameter configuration. The final base model will then be fitted to the entire training set using these optimal hyperparameters.

Final Validation

To assess the incremental predictive value of MET, a copy of the final base model will be created with the addition of the MET predictor to create an extended model. Both the base and the extended models will be fitted on the complete training set and evaluated on the held-out test set, which is used exclusively at this stage and remains untouched throughout all preceding model development steps.

The incremental predictive value of MET will be assessed using C-index, calibration and decision curve analysis (DCA), comparing the base and extended models on the held-out test set. Clinical utility will be evaluated by comparing net benefit between the base and extended models at pre-specified time horizons of 30, 90 and 365 days, with net benefit plotted across a range of clinically plausible threshold

probabilities, defined as 0 to 2% for 30-day mortality 0 to 5% for 90-day mortality, and 0 to 15 % for 365-day mortality, reflecting the expected event rates at each time horizon.

5.4 Sensitivity analysis

1. If the proportion of missing values exceeds 10% in any predictor used for model development, the internal validation step of the primary analysis will be repeated as a sensitivity analysis using multiple imputation by chained equations. Imputation models will be estimated exclusively on training data and applied to test data to prevent leakage. Performance estimates will be pooled across imputed datasets using Rubin's rules.

2. The primary analysis will be repeated from the step where the final model type is decided, with an alternative set of predictors: age, sex and RCRI.

5.5 P-values

A two-sided significance threshold of $\alpha = 0.05$ will be applied throughout. Results will be reported in accordance with the TRIPOD+AI statement.

5.6 Missing Data

The pattern and extent of missing data will be characterized as part of the descriptive analysis, with a table reporting the proportion of missing values for each covariate and outcome variable. The plausibility of the missing data mechanism (missing completely at random, at random, or not at random) will be assessed to inform the missing data handling strategy applied.

6. Preregistration and Study Reporting

6.1 Preregistration

The study protocol and statistical analysis plan will be prospectively registered on ClinicalTrials.gov prior to data analysis. The assigned registration number (NCT07603622) will be reported in all publications and presentations arising from this study. Any amendments to the registered protocol or statistical analysis plan made after data analysis has commenced will be documented, dated, and reported transparently as post-hoc modifications in all disseminated outputs.

6.2 Study Reporting

This study will be reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis **+** Artificial Intelligence (TRIPOD + AI) statement.

References

1. Halvorsen S, Mehilli J, Cassese S, et al. 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non-cardiac surgery. *European heart journal*. 2022;43(39). doi:10.1093/EURHEARTJ/EHAC270

2. Valadkhani A, Sebghati S, Piehl J, Bell M. The Impact of Preoperative Functional Capacity on Postoperative Mortality and Morbidity. A Prospective Cohort Study. *Anesthesiology*. doi:10.1097/ALN.0000000000005779