



# **The Maximizing Extubation outcomes Through Educational and Organizational Research Trial**

Statistical Analysis Plan version 1.0 – April 27, 2026

## TABLE OF CONTENTS

<b>1</b>	<b>Administrative information.....</b>	<b>4</b>
1.1	Title and trial registration.....	4
1.2	Statistical analysis plan (SAP) version .....	4
1.3	Protocol version .....	4
1.4	SAP revisions .....	4
1.5	Roles and responsibility .....	4
<b>2</b>	<b>Introduction .....</b>	<b>4</b>
2.1	Background and rationale .....	4
2.2	Objectives.....	5
2.3	Hypotheses .....	6
<b>3</b>	<b>Study methods .....</b>	<b>7</b>
3.1	Trial design .....	7
3.2	Randomization .....	7
3.3	Sample size .....	8
3.4	Framework .....	9
3.5	Statistical interim analyses and stopping guidance .....	9
3.6	Timing of final analysis .....	10
3.7	Timing of outcome assessments .....	10
<b>4</b>	<b>Statistical principles.....</b>	<b>10</b>
4.1	Confidence intervals and P values.....	10
4.2	Adherence .....	10
4.3	Analysis populations.....	11
<b>5</b>	<b>Trial population .....</b>	<b>11</b>
5.1	Screening data.....	11
5.2	Eligibility .....	11
5.3	Recruitment.....	11
5.4	Withdrawal/follow-up.....	11
5.5	Baseline subject characteristics .....	12
<b>6</b>	<b>Analysis.....</b>	<b>12</b>
6.1	Outcomes definitions .....	12
6.2	Analysis methods .....	13
6.3	Missing data .....	15
6.4	Additional analyses .....	15
6.5	Statistical software .....	16
<b>7</b>	<b>References .....</b>	<b>17</b>

## LIST OF ABBREVIATIONS

aHR	Adjusted hazard ratio
aRR	Adjusted rate ratio
BMI	Body mass index
CFR	Code of Federal Regulations
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
EHR	Electronic health record
GLMM	Generalized linear mixed model
HFNC	High-flow nasal cannula oxygen
HR	Hazard ratio
ICC	Intraclass correlation
ICU	Intensive care unit
IPE	Interprofessional education
ITT	Intention-to-treat
MAR	Missing at random
METEOR	Maximizing Extubation Outcomes Through Educational and Organizational Research
MICE	Multiple imputation by chained equations
NIV	Noninvasive ventilation
NDI	National Death Index
OE	Online education
OHRP	Office for Human Research Protections
P	Protocol
PH	Proportional hazards
RCT	Randomized controlled trial
RR	Relative risk
SAP	Statistical analysis plan
SOFA	Sequential Organ Failure Assessment
UC	Usual care
VAE	Ventilator-associated event

## **1 ADMINISTRATIVE INFORMATION**

### **1.1 Title and trial registration**

The Maximizing Extubation outcomes Through Educational and Organizational Research (METEOR) Trial was registered at ClinicalTrials.gov on August 31, 2022 (NCT05523479).

### **1.2 Statistical analysis plan (SAP) version**

The current version of the SAP (version 1.0) was finalized on April 27, 2026.

### **1.3 Protocol version**

This SAP refers to version 1.6 (March 7, 2025) of the protocol.

### **1.4 SAP revisions**

Version 1.0 of the SAP was approved by the Steering Committee on April 27, 2026, and by the Data and Safety Monitoring Board on April 30, 2026.

### **1.5 Roles and responsibility**

The following investigators contributed to the SAP:

Timothy D Girard, MD, MSCI, University of Pittsburgh, Principal Investigator  
(Joyce) Chung-Chou H. Chang, PhD, University of Pittsburgh, Senior Statistician  
Niall T Prendergast, MD, University of Pittsburgh, Co-Investigator  
Jeremy M Kahn, MD, MS, University of Pittsburgh, Co-Investigator

## **2 INTRODUCTION**

### **2.1 Background and rationale**

Nearly one million patients require invasive mechanical ventilation for acute respiratory failure in the United States each year. Most of these patients will recover to the point of extubation, yet even those who are extubated remain vulnerable to complications and poor outcomes. Multiple high-profile randomized controlled trials have shown that two preventive post-extubation respiratory therapies-noninvasive ventilation (NIV) and high-flow nasal cannula oxygen (HFNC)-can prevent recurrent respiratory failure, reintubation, and death in this population. Despite this evidence, however, these therapies remain severely underutilized, leading to preventable morbidity and mortality. To address this implementation gap, the investigators will conduct the Maximizing Extubation outcomes Through Educational and Organizational Research (METEOR) Trial, a cluster-randomized, stepped-wedge, type 2 hybrid effectiveness-implementation trial of interprofessional education (IPE) about preventive post-extubation NIV and HFNC with and without clinical protocols.

The METEOR Trial was designed based on extensive preliminary studies, during which the investigators identified barriers to adoption of preventive post-extubation respiratory care and pilot tested IPE as an implementation strategy in the intensive care unit (ICU). These studies

revealed that a major barrier to implementation is the lack of a shared understanding about the value of these therapies within the interprofessional ICU team; a theory-based IPE intervention designed to create a shared understanding and support "transactive memory" among team members is both feasible and acceptable; and IPE can be strengthened by linking it with a clinical protocol. During the METEOR Trial, the investigators will randomize ICUs to one of four implementation strategies: an active control, protocol-directed care, IPE, or a combination of protocol-directed care and IPE. In parallel, the investigators will randomize ICUs to one of two clinical strategies, one emphasizing either post-extubation NIV or HFNC based on patient risk vs. one emphasizing post-extubation HFNC for all patients.

## 2.2 Objectives

The objectives listed below expand upon those pre-specified in the protocol to provide the additional specificity required for the SAP; the hypotheses in section 2.3, alternatively, are reproduced as published in the protocol.

The primary **implementation objective** is to test the effectiveness of IPE, with or without a clinical protocol, on the implementation of two preventive, post-extubation respiratory support therapies (NIV and HFNC). Because implementation interventions will be deployed sequentially during a batched stepped wedge cluster randomized trial, we will evaluate their effects by comparing implementation and clinical endpoints across randomized intervention phases. In this context, the more specific implementation objectives are to:

- Determine whether traditional online education (OE) improves the primary implementation endpoint compared with usual care (UC).
  - If the adjusted risk difference between OE and UC exceeds 5% and the 95% CI excludes 0%, we will use OE as the primary comparator for subsequent analyses.
  - If the adjusted risk difference between OE and UC is  $\leq 5\%$  or the 95% CI includes 0%, we will combine OE and UC and use this combination as the primary comparator.
- Determine whether interprofessional education (IPE), when delivered individually or with a clinical protocol (P), improves implementation endpoints relative to the primary comparator.
- Determine whether the combined implementation strategy (IPE+P) produces greater implementation success than either IPE alone or P alone and whether the effects of IPE and P are additive or synergistic.

The primary **clinical objective** is to compare the effectiveness of two preventive, post-extubation respiratory support therapies (NIV and HFNC) with that of conventional post-extubation oxygen on patient-centered clinical endpoints. Because the clinical effectiveness of the preventive, post-extubation respiratory support therapies (NIV and HFNC) depend directly on successful implementation of these therapies, the analyses examining clinical endpoints are explicitly linked to implementation success and take into account implementation of the respiratory support therapies. Accordingly, the more specific effectiveness objectives are to:

- Estimate the effects of the implementation interventions (OE, IPE alone, P alone, and IPE+P) on clinical endpoints using the empirically defined primary comparator (OE alone or UC+OE), during which conventional oxygen is expected to be the predominant post-extubation respiratory support therapy.
- Evaluate the effects of the following post-extubation respiratory support strategies during the full implementation phase (IPE+P), when implementation is expected to be maximized, examining their effects on clinical endpoints relative to the empirically defined primary comparator phase, when conventional oxygen is expected to be the predominant post-extubation respiratory support therapy, and relative to each other.
  - Post-extubation NIV for high-risk patients and HFNC for low-risk patients
  - Post-extubation HFNC for all patients, regardless of risk
- Determine whether partial implementation (i.e., of IPE alone or P alone) is sufficient to confer clinical benefit if the combined implementation strategy (IPE+P) does not result in superior clinical effectiveness.

Together, these implementation and clinical objectives ensure that the results of clinical effectiveness analyses are interpretable, are aligned with observed implementation fidelity, and are relevant to real-world ICU practice.

## 2.3 Hypotheses

This type 2 hybrid implementation-effectiveness trial will test a set of pre-specified, interrelated hypotheses about both implementation success and clinical effectiveness. The implementation and clinical hypotheses are related but not symmetric. Confirmation of the primary implementation hypotheses is not contingent on confirmation of the clinical hypothesis; i.e., implementation success is interpretable on its own merits regardless of clinical effectiveness. The clinical hypothesis, however, is interpreted conditional on implementation success, specifically on evidence that the implementation interventions increased use of the preventive respiratory support therapies.

The following **implementation hypotheses** focus on the effects of three ICU-level implementation interventions (described in protocol section 6.1) on the implementation of two preventive, post-extubation respiratory support therapies (NIV and HFNC) and expand upon hypotheses listed in protocol section 2.2.

- **Implementation hypothesis 1** (primary): IPE is superior to the empirically defined primary comparator as an implementation strategy in the ICU.
- **Implementation hypothesis 2** (secondary): The benefits of IPE are increased when IPE is paired with a clinical protocol.

The following **clinical hypotheses** focus on the effects of ICU-level implementation interventions (described in protocol section 6.1) and their content (described in protocol section 6.2) on patient-centered clinical endpoints.

- **Clinical hypothesis:** Preventive post-extubation NIV for high-risk patients and preventive post-extubation HFNC for low-risk patients are both superior to current clinical practice (i.e., conventional post-extubation oxygen therapy).

## 3 STUDY METHODS

### 3.1 Trial design

The METEOR Trial is a batched stepped wedge cluster randomized type 2 hybrid implementation-effectiveness trial comparing three implementation interventions—traditional online education, interprofessional education, and an evidence-based clinical protocol—with the latter two being examined both alone and in combination. The trial also compares two evidence-based approaches to preventive post-extubation respiratory care (post-extubation HFNC for all eligible patients and risk-stratified post-extubation NIV or HFNC) with conventional post-extubation oxygen therapy and with each other. The implementation interventions and content are described in detail in the protocol.

Clusters (ICUs or groups of ICUs) are randomized rather than individuals, and each cluster will cross over from control to intervention at a randomized timepoint. This design is ideal for the METEOR Trial for three reasons. First, patient-level randomization is scientifically problematic due to a high risk of contamination of the control group by a systems-level intervention.<sup>1</sup> Second, a parallel cluster RCT is infeasible because ICUs will be unwilling to serve as controls during the entire trial period.<sup>2</sup> Third, recruitment of all ICUs at the same time may significantly burden a health system that is severely constrained.<sup>3</sup>

We will employ three variations on the traditional stepped wedge design:

1. A factorial design will allow us to estimate the effects of two implementation strategies (IPE and a clinical protocol) separately as well as together.<sup>4</sup>
2. A concurrent design will allow us to compare two post-extubation strategies (HFNC for all patients vs. NIV for high-risk patients and HFNC for low-risk patients) with conventional post-extubation oxygen therapy and with each other.<sup>4</sup>
3. A batched design allows for the recruitment of clusters (i.e., ICUs) throughout the duration of the trial rather than requiring that all clusters commence participation in the trial at the same time.<sup>3</sup>

### 3.2 Randomization

The unit of randomization is the cluster, which may consist of one or more ICUs.<sup>3</sup> In the remainder of this SAP, we often use “ICU” as shorthand for cluster. At baseline, participating ICUs are receiving usual care, i.e., no structured education regarding post-extubation respiratory support therapies. As shown in [Table 1](#), a batch will include between two and six clusters, which will be randomly assigned using a computer-generated randomization scheme to one of six different sequences for crossing over from usual care (UC, a passive baseline control) to online education (OE, an active control) to interprofessional education (IPE) or

protocol (P) individually, then finally to both IPE and P. In batch 1, the ICU randomly assigned to sequence 1 will continue UC during the first step of the trial (S1), with a step lasting one month, then will deploy OE for three steps, then IPE alone for two steps before finally deploying both IPE and P together. Alternatively, the ICU assigned to sequence 2 will continue UC for one step, then deploy OE for four steps, then deploy P alone for two steps, before finally deploying both IPE and P together. Sequences 3 and 4 are the same as sequences 1 and 2, respectively, but crossover to OE from UC occurs later, as shown. Batch 2 will include the same six sequences but will begin crossover six steps after Batch 1. Ultimately, all participating ICUs (not all of which are shown in [Table 1](#)) will deploy IPE plus P intervention by the end of the trial.

**Table 1. Stepped-wedge deployment schedule**

Batch	Sequence	ICU	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S...
1	1	1	UC	OE	OE	OE	IPE	IPE	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
1	2	2	UC	OE	OE	OE	OE	P	P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
1	3	3	UC	UC	UC	OE	OE	OE	IPE	IPE	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
1	4	4	UC	UC	UC	OE	OE	OE	P	P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
1	5	5	UC	UC	UC	UC	UC	OE	OE	OE	IPE	IPE	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
1	6	6	UC	UC	UC	UC	UC	OE	OE	OE	OE	P	P	IPE+P	IPE+P	IPE+P	IPE+P	IPE+P
2	1	7							UC	OE	OE	OE	IPE	IPE	IPE+P	IPE+P	IPE+P	IPE+P
2	2	8							UC	OE	OE	OE	OE	P	P	IPE+P	IPE+P	IPE+P
2	3	9							UC	UC	UC	OE	OE	OE	IPE	IPE	IPE+P	IPE+P
2	4	10							UC	UC	UC	OE	OE	OE	OE	P	P	IPE+P
2	5	11							UC	UC	UC	UC	UC	OE	OE	OE	IPE	IPE
2	6	12							UC	UC	UC	UC	UC	OE	OE	OE	OE	P
3	1	13													UC	OE	OE	OE
3	2	14													UC	OE	OE	OE
...	...	...													UC	UC	UC	OE

In addition to being randomly assigned to a specific deployment sequence, each cluster will also be randomly assigned to one evidence-based strategy promoted by the IPE and P—a strategy emphasizing post-extubation HFNC for all patients vs. a strategy emphasizing post-extubation NIV for high-risk patients and HFNC for low-risk patients. Because the assigned implementation content is not deployed during the UC phase and is dependent on the effectiveness of the implementation strategies (OE, IPE, P, and IPE+P), the analyses examining the clinical effectiveness of different implementation content will be defined based on the results of the analyses examining implementation endpoints.

### 3.3 Sample size

To inform sample size and power calculations, we queried the UPMC ICU registry for data from 28 UPMC ICUs expected to participate. The rate of post-extubation NIV or HFNC among eligible patients was 2.3% with an intraclass correlation (ICC) of 0.418. The in-hospital mortality rate was 21.7% with an ICC of 0.064. The number of eligible patients averaged 30 per month per ICU. In power calculations, we conservatively assumed an average of 23.3 patients with complete assessments per cluster per one-month time period. We originally planned to include 28 clusters but later (with DSMB approval) revise the sample size to 24 clusters and updated the power calculations. Based on our batched cluster-randomized, stepped-wedge design (see [Table 1](#)) and a 5% type I error rate, we estimate that the METEOR Trial will have 90% power to

detect a 1.44% increase in the primary implementation endpoint and a 4.56% decrease in in-hospital mortality for key comparisons (e.g., IPE+P vs the primary comparator). These power calculations reflect per-hypothesis power under the hierarchical testing framework. The overall probability of confirming multiple hypotheses depends on the correlation between endpoints and the sequential nature of testing. Given the conditional structure of the trial, we consider these per-hypothesis power estimates adequate to inform the trial design and interpretation.

### 3.4 Framework

All analyses will be conducted using a superiority hypothesis testing framework. For the primary analyses, we will use a hierarchical testing framework that reflects the conditional relationship between the primary implementation and clinical hypotheses. We will test implementation hypothesis 1 first. If implementation hypothesis 1 is confirmed, we will then interpret the analysis testing the clinical hypothesis as confirmatory. Within each hypothesis, the intersection-union principle applies—implementation hypothesis 1 requires that IPE is superior to the primary comparator, and the clinical hypothesis requires that both implementation contents are superior to current clinical practice.

### 3.5 Statistical interim analyses and stopping guidance

As pre-specified in the protocol (section 7.3), we conducted an interim analysis after approximately 18 months of recruitment (halfway through the trial) to monitor for evidence that one of the two implementation contents resulted in greater effectiveness than the other. We anticipated that both implementation contents would reduce mortality compared with conventional post-extubation oxygen therapy. Because one implementation content group could experience greater benefit than the other, we conducted the interim analysis using a two-sided significance test. We considered the interim analysis to be exempt from the hierarchical requirement because it was safety analysis rather than a confirmatory one. The hypotheses tested in the interim analyses were:

- **H<sub>0</sub>:** The difference in in-hospital mortality for those randomized to preventive post-extubation NIV for high-risk patients and preventive post-extubation HFNC for low-risk patients vs those randomized to preventive post-extubation HFNC is  $\leq 5\%$ .
- **H<sub>1</sub>:** The difference in in-hospital mortality for those randomized to preventive post-extubation NIV for high-risk patients and preventive post-extubation HFNC for low-risk patients vs those randomized to preventive post-extubation HFNC is  $>5\%$ .

We fit a generalized linear mixed model (GLMM) with a log link and binomial error distribution for in-hospital mortality. Fixed effects included the comparison arm (control vs. intervention), time (month), and other covariates for adjustment. We included a random intercept for clusters to account for within-ICU correlations. We conducted a two-sided test with an overall type I error rate of 5%. Using the O'Brien-Fleming method to control for an overall type I error rate of 0.05 and accounting for one interim analysis after approximately 50% recruitment, we planned to modify the trial if  $p < 0.0055746$  at the interim analysis, but this threshold was not met.

### **3.6 Timing of final analysis**

All endpoints will be analyzed collectively after the last cluster has completed the sequence determined by randomization.

### **3.7 Timing of outcome assessments**

All clinical endpoints except 90-day mortality will be recorded in the EHR during each subject's index hospitalization. For subjects whose hospitalization is shorter than 90 days, we will collect 90-day vital status using the National Death Index (NDI).

The primary implementation endpoint will be recorded in the EHR during each subject's index hospitalization. Secondary implementation endpoints will be obtained by an ancillary survey study and are described in detail in a separate SAP.

## **4 STATISTICAL PRINCIPLES**

### **4.1 Confidence intervals and P values**

We will use a two-sided overall significance level of 5% and report 95% confidence intervals (CIs). Although this trial involves multiple pre-specified hypotheses and endpoints, we will make no adjustments for multiple comparisons. As described in section 3.4, we will use a hierarchical testing framework for the primary analyses. Under this hierarchical framework, the type I error rate is controlled at the nominal level without multiplicity adjustment.

For secondary endpoints, we follow standard recommendations and practice in confirmatory clinical trials, whereby pre-specified secondary endpoints are analyzed without adjustment for multiple comparisons. Each comparison is hypothesis-driven and grounded in clinical and implementation plausibility rather than being exploratory or hypothesis-generating in nature. Accordingly, we will interpret secondary endpoint results in the context of the totality of evidence and with appropriate consideration of the number of comparisons made, rather than as independently confirmatory findings.

Certain pre-specified comparisons (e.g., OE vs UC) serve an interpretive function rather than testing a standalone hypothesis. We will report these results descriptively along with results of tests they inform. Lastly, we will interpret the results of sensitivity analyses and heterogeneity of treatment effect analyses as exploratory and will not adjust for multiplicity.

### **4.2 Adherence**

Since the implementation interventions are education, we will define adherence as participation in the education and non-adherence as lack of participation. As described in section 6.1, we will report and compare the number of providers who completed an implementation intervention as a secondary implementation endpoint. We will also report the number of providers invited to participate and calculate adherence rates by dividing the number of providers who completed each implementation intervention by the number invited.

### **4.3 Analysis populations**

We will conduct the primary analyses according to a cluster-period intention-to-treat (ITT) principle, analyzing each cluster-period cell according to its randomized assignment and implementation phase, irrespective of actual intervention fidelity. We will analyze all eligible patients according to the ICU in which their extubation occurred. Because the implementation interventions target post-extubation respiratory care, we will attribute patients who are intubated in one ICU and later extubated in a different ICU to the ICU where extubation occurred for all primary analyses.

Because the clinical effectiveness of the preventive, post-extubation respiratory support therapies (NIV and HFNC) depend directly on successful implementation of these therapies, for the analyses comparing the two implementation contents (post-extubation NIV for high-risk patients and HFNC for low-risk patients vs. post-extubation HFNC for all patients, regardless of risk), we will define the ITT population as all eligible patients who are extubated during periods in which the cluster has transitioned to the full IPE+P implementation phase, and we will analyze patients according to the implementation content assigned to their cluster, irrespective of actual treatment delivery.

## **5 TRIAL POPULATION**

### **5.1 Screening data**

Because the trial meets the Office for Human Research Protections (OHRP) regulations regarding waiver of informed consent (45 CFR 46.116(f)(3)), we will include all eligible subjects. We will track the number of mechanically ventilated patients treated in participating ICUs during the trial and will report the number of these patients who meet eligibility criteria (and are thus included as subjects) and the number who are not included because they do not meet eligibility criteria.

### **5.2 Eligibility**

All adults treated with invasive mechanical ventilation >24 hours in participating ICUs will be included. This pragmatic trial has no exclusion criteria. Having no restrictions on age, comorbidities, or expectations regarding outcomes reflects the realities of real-world decision-making.

### **5.3 Recruitment**

In the CONSORT flow diagram, we will report the number of mechanically ventilated patients treated in participating ICUs during the trial, the number of these patients who are included (i.e., recruited) as subjects, the number who are not included (i.e., who do not meet eligibility criteria), and the reasons they are not included.

### **5.4 Withdrawal/follow-up**

Because this pragmatic trial involves hospitalized patients with waiver of the requirement for informed consent, we anticipate no withdrawals. Additionally, since nearly all clinical endpoints

are recorded in the EHR, we anticipate very little loss to follow-up. Some endpoint data, however, may be lost due to transfer out of participating hospitals. We will report the number of subjects who are lost to follow-up and describe the percentage and reasons for missing data.

## 5.5 Baseline subject characteristics

We will summarize the following baseline characteristics, presenting results as median [interquartile range] for continuous variables and n (%) for categorical variables: age, sex, body mass index (BMI), Elixhauser Comorbidity Index,<sup>5</sup> Sequential Organ Failure Assessment (SOFA) score<sup>6,7</sup> at ICU admission, primary admission diagnosis, chronic heart disease, and chronic lung disease.

## 6 ANALYSIS

### 6.1 Outcomes definitions

This type 2 hybrid effectiveness-implementation trial will measure endpoints that reflect the effectiveness of the interventions on both implementation and clinical endpoints.

#### Implementation endpoints

- **Primary implementation endpoint:** Rate of use of post-extubation NIV or HFNC among eligible subjects throughout the trial. Eligible subjects are those who are extubated (i.e., their endotracheal tube is removed) after >24 hours of mechanical ventilation. Subjects with a tracheostomy or with “comfort measures only” orders at the time of extubation are not eligible for post-extubation NIV or HFNC.
- **Secondary implementation endpoints:**
  - a. Number of providers who completed an implementation intervention
  - b. Number of eligible subjects receiving care from providers who completed an implementation intervention
  - c. Use of post-extubation NIV or HFNC among eligible subjects 6 months after the implementation intervention (IPE plus protocol) is fully deployed

#### Clinical endpoints

- **Primary clinical endpoint:** In-hospital mortality truncated at 60 days from intubation
- **Secondary clinical endpoints:**
  - a. 90-day survival
  - b. ICU length of stay
  - c. Hospital length of stay
  - d. Rate of post-extubation respiratory failure (defined as reintubation within 48 hours of planned extubation)
  - e. Time to successful liberation from mechanical ventilation (defined as breathing without assistance for at least 48 hours)

- f. 28-day ventilator-free days (defined as the number of days alive and breathing without assistance during the 28 days after index intubation with death equated to 0 ventilator-free days)
- g. Rate of ventilator-associated events (VAEs)
- h. Organ failure (measured using daily Sequential Organ Failure Assessment score)

## 6.2 Analysis methods

We will use analytical approaches that account for the batched stepped-wedge design and will conduct separate analyses examining implementation and clinical endpoints. As described in section 4.3, all primary analyses will follow a cluster-period ITT principle, with each cluster-period cell analyzed according to its randomized assignment and implementation phase and each eligible patient analyzed according to the ICU in which extubation occurs.

Specifically, in all models we will:

- Adjust for calendar time (or step) to account for secular trends, allowing separate time trends to be estimated for intervention and control conditions;
- Include cluster-level random intercepts to account for within-cluster correlation;
- Incorporate cluster-by-time random effects or interactions, where appropriate, to allow for heterogeneity in secular trends across clusters;
- Adjust for baseline patient characteristics that influence mortality, including age, Elixhauser Comorbidity Index, chronic heart disease, chronic lung disease, BMI, and SOFA score at ICU admission;
- Adjust for two cluster-level variables: academic vs. community hospital and annualized ICU volume (mean number of patients receiving mechanical ventilation per year during the trial); and
- Use robust variance estimation and Kenward and Roger small sample correction to reduce potential inflation of the type-I error rate due to a small number of clusters.

We define model specifications herein and apply them consistently across endpoints within each endpoint class. We will include the above-mentioned terms to account for secular trends and heterogeneity across clusters.

### Implementation endpoint analyses

As specified in Section 2.2, we will first determine whether OE improved the primary implementation endpoint compared with UC, and the results of this analysis will determine the primary comparator that we will use consistently in all analyses of implementation endpoints. If the adjusted risk difference between OE and UC exceeds 5% and the 95% CI excludes 0%, we will use OE as the primary comparator for subsequent analyses. Alternatively, if the adjusted risk difference between OE and UC is  $\leq 5\%$  or the 95% CI includes 0%, we will combine OE and UC and use this combination as the primary comparator.

We will use a GLMM as the primary model examining effects on implementation endpoints. We will fit the model to estimate the effects of cluster-level implementation strategies on the primary implementation endpoints and report adjusted effect estimates with 95% CIs. The model will include indicator variables for the following implementation interventions:

- Primary comparator condition (OE alone or UC+OE, as previously defined)
- IPE alone
- P alone
- IPE+P (the combined implementation strategy)

We will specify the model to yield time- and cluster-adjusted estimands listed below. The primary confirmatory estimand is IPE vs. primary comparator (implementation hypothesis 1). The remaining estimands are specified to test secondary implementation hypotheses and to provide context for their interpretation:

- IPE vs. primary comparator
- P vs. primary comparator
- IPE+P vs. primary comparator
- IPE+P vs. IPE alone and IPE+P vs. P alone

To assess whether the effects of IPE and P are additive or synergistic, we will include an interaction term between IPE and P in all models. All pre-specified terms, including this interaction term, will be retained in the model regardless of statistical significance, in keeping with the confirmatory nature of this trial and to avoid post-hoc model selection.

### **Clinical effectiveness analyses**

After determining the effects of implementation interventions on implementation endpoints, we will examine their effects on patient-centered clinical endpoints, conditional on observed implementation effects (as described in Section 2.2). To examine the effects of implementation interventions on clinical endpoints, we will fit models analogous to the implementation analyses models, with the implementation interventions (primary comparator, IPE alone, P alone, and IPE+P) as the primary independent variables.

We will select appropriate models based on the endpoint type:

- Binary endpoints (e.g., in-hospital mortality) will be analyzed using GLMMs with appropriate error distributions and link functions.
- Time-to-event endpoints will be analyzed using Cox proportional hazards (PH) models with a shared frailty term to account for clustering.
- Continuous endpoints will be analyzed using linear-mixed effects models.

We will report effect estimates as adjusted rate ratios (aRRs), adjusted hazard ratios (aHRs), or adjusted mean differences, as appropriate, with 95% CIs. We will use model diagnostics to assess the appropriateness of assumptions. In the case of non-convergence of the GLMM, we

will use alternative estimation strategies, including reducing the number of integration points or using a Laplace approximation.

As explained in section 4.3, for the analyses comparing the two implementation contents (post-extubation NIV for high-risk patients and HFNC for low-risk patients vs. post-extubation HFNC for all patients, regardless of risk), we will analyze only eligible patients who were extubated during periods in which the cluster had transitioned to the full IPE+P implementation phase, and we will analyze patients according to the implementation content assigned to their cluster, irrespective of actual treatment delivery.

### **6.3 Missing data**

Given the pragmatic design of the trial and the use of routinely recorded EHR data, we anticipate a low proportion of missing data for the primary implementation and clinical endpoints. We will summarize the extent and patterns of missing data for all endpoints and key covariates prior to analysis.

The primary analytic models are mixed-effects models estimated via maximum likelihood, which incorporate all available outcome data from all randomized clusters and participants without requiring complete cases or imputation. These models yield valid inference under a missing at random (MAR) assumption, conditional on covariates and random effects included in the model. We will therefore fit primary models using all available data rather than restricting to complete cases.

For missing covariates exceeding 5% missingness, we will use multiple imputation by chained equations (MICE), generating 20 imputed datasets. Each dataset will be analyzed separately using the primary analytic models, and results will be combined across imputed datasets using Rubin's rules.

### **6.4 Additional analyses**

In addition to the main analyses specified in section 6.2, we will conduct sensitivity analyses, heterogeneity of treatment effect analyses, a path analysis, and a per-protocol analysis.

#### **Sensitivity analyses**

Because the MAR assumption is untestable, we will assess the robustness of our primary analyses findings to departures from MAR using a pattern mixture modeling approach, applying a range of delta adjustments to evaluate how strongly the MNAR mechanism would need to operate to materially alter inference. We will also repeat all analyses using only the first episode of mechanical ventilation to determine if the results are heavily influenced by the subset of subjects who experienced more than one episode of mechanical ventilation during the trial period.

## Heterogeneity of treatment effects

We will determine whether obesity modifies the effects of the preventive, post-extubation respiratory support therapies (NIV and HFNC) by including an interaction term in the model analyzing the effects of these therapies on clinical endpoints. We will analyze BMI as a continuous, nonlinear term. We will report the p value for the interaction and will also display in a Forest plot the point estimates and 95% CIs of treatment effects stratified by World Health Organization BMI categories (underweight, normal, overweight, and obese).

We will determine whether hospital type (academic vs. community) modifies the effects of the implementation interventions by including an interaction term in the model analyzing the effects of these interventions on implementation endpoints. We will analyze hospital type as a dichotomous variable defined by the presence (academic) or absence (community) of physicians in postgraduate training (residents and/or fellows). We will report the p value for the interaction and will also display in a Forest plot the point estimates and 95% CIs of treatment effects stratified by hospital type.

## Path analyses

We will also perform a path analysis linking our implementation outcomes to our clinical outcomes with a goal of determining the degree to which process change mediates our observed clinical effects.<sup>8</sup>

## Per-protocol analyses

Finally, we will conduct an exploratory per-protocol analyses to assess the robustness of the primary ITT analysis results and to evaluate the potential impact of delayed or incomplete deployment of implementation interventions. In these analyses, we will classify patients according to the dates of their extubations relative to the actual availability of the implementation interventions (Table 2). In these per-protocol analyses, we will use the same modeling framework, covariates, and outcomes as in the primary ITT analyses. We will interpret the results of the per-protocol analyses in relation to the ITT analysis results and will not use the per-protocol results to redefine the primary estimands or conclusions.

Table 2. Per-protocol definitions	
Intervention	Intervention deployment date
Usual care	Date cluster entered trial
Online education	Date of first email inviting ICU providers in a cluster to complete online education
Interprofessional education	Date of first interprofessional education workshop in a cluster
Clinical protocol	Date clinical protocol template was provided to local champion in a cluster

## 6.5 Statistical software

We will use Stata version 19 or newer (College Station, TX, USA) and R version 4.4.0 or newer (The R Project) for all analyses.

## 7 REFERENCES

1. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655. PubMed Central PMCID: PMC2769032.
2. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* 2011;64:936–48.
3. Kasza J, Bowden R, Hooper R, Forbes AB. The batched stepped wedge design: A design robust to delays in cluster recruitment. *Statistics in medicine* 2022;41:3627–41.
4. Lyons VH, Li L, Hughes JP, Rowhani-Rahbar A. Proposed variations of the stepped-wedge design can be used to accommodate multiple interventions. *J Clin Epidemiol* 2017;86:160–7. PubMed Central PMCID: PMC5835387.
5. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
6. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22:707–10.
7. Vincent JL, Mendonca Ad, Cantraine F, Moreno R, Takala J, Suter P, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. *Crit Care Med* 1998;26:1793–800.
8. Edwards JR, Lambert LS. Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychol Methods* 2007;12:1–22.