

Framework for Patient-Device Model-Based In Silico Clinical Trials in Mechanical Ventilation: A Case Study

Carlotta Hennigs

`carlotta.hennigs@uni-luebeck.de`

University of Lübeck

Georg Männel

University of Lübeck

Franziska Bilda

University of Lübeck

Helene Selpien

University Hospital Schleswig-Holstein

Dirk Schädler

University Hospital Schleswig-Holstein

Michael Imhoff

qtec Services GmbH

Philipp Rostalski

University of Lübeck

Article

Keywords: Mechanical ventilation, in silico clinical trial, patient-device model, weaning

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10840046/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Framework for Patient-Device Model-Based In Silico Clinical Trials in Mechanical Ventilation: A Case Study

Carlotta Hennigs^{1,2*}, Georg Männel^{3,2}, Franziska Bilda^{1,2},
Helene Selpien⁴, Dirk Schädler⁴, Michael Imhoff^{5,6},
Philipp Rostalski^{1,2}

^{1*}Institute of Electrical Engineering in Medicine, University of Luebeck,
Lübeck, Germany.

²Fraunhofer Research Institution for Individualized Medical Technology
and Engineering IMTE, Lübeck, Germany.

³Institute of Robotics and Cognitive Systems, University of Luebeck,
Lübeck, Germany.

⁴Department of Anesthesiology and Intensive Care Medicine University
Hospital Schleswig-Holstein, Kiel, Germany.

⁵qtec Services GmbH, Lübeck, Germany.

⁶Department of Medical Informatics, Biometrics, and Epidemiology
Ruhr-Universität Bochum, Bochum, Germany.

*Corresponding author(s). E-mail(s): carlotta.hennigs@uni-luebeck.de;

Abstract

Purpose: This work addresses the need for a standardized, credibility-aware in silico clinical trial (ISCT) framework to support the preclinical evaluation of automated ventilator functions—including emerging multi-variable, adaptive, and AI-driven control strategies—within regulatory frameworks such as ASME V&V 40 and FDA guidance.

Methods: We propose a device-agnostic ISCT workflow that follows a top-down, risk-informed design logic: starting from a clearly defined question of interest and context of use, through model definition, virtual cohort and scenario generation, simulation execution, and quantitative analysis. The framework integrates verification and validation activities across all workflow steps. Its utility is demonstrated through a case study evaluating an automated weaning protocol across a virtual cohort of 400 adult ICU patients, spanning representative and edge-case

strata (ARDS, COPD, postoperative, neuromuscular weakness, cardiogenic pulmonary edema). Each patient is simulated in nine perturbation scenarios probing changes in mechanics, drive, and gas exchange.

Results: The automated weaning protocol maintains acceptable working ranges for tidal volume, respiratory rate, and end-tidal carbon dioxide across most scenarios, while lung-protective targets are challenged under severe gas exchange impairment. Patient-ventilator asynchrony and inspiratory pressure oscillations occur in a scenario- and phenotype-dependent manner, with oscillation frequencies (11.1%) comparable to clinical observations (15.5%). Influence analysis identifies disease state, sedation level, and initial inspiratory pressure as dominant performance determinants.

Conclusion: The framework provides a reusable, regulatory-aligned pathway for systematic stress-testing and failure-mode identification of automated ventilation strategies prior to clinical investigation. It enables transparent, decision-relevant evidence generation under real-world complexity, thereby supporting iterative design, regulatory submission, and robust clinical translation of automated ventilator functions.

Keywords: Mechanical ventilation, in silico clinical trial, patient-device model, weaning

1 Introduction

Medical devices are developed and evaluated under increasing ethical, economic, and logistical constraints. Conventional preclinical and clinical studies are expensive, time-consuming, and can cover only a fraction of clinically relevant conditions. Rare but high-consequence events, complex interactions between device behaviour, patient physiology, and clinical workflows are therefore often under-represented in the available evidence. At the same time, modern devices increasingly incorporate software-driven, adaptive, and physiological closed-loop functions - moving beyond traditional single-variable feedback controllers (which have been established in ventilation and other therapy devices for decades) toward multi-variable, patient-adaptive systems as envisioned by IEC 60601-1-10 - which challenge verification, validation, and regulatory assessment paradigms that were originally designed for comparatively static technologies. The EU Medical Device Regulation (MDR 2017/745) fundamentally prioritizes patient safety by establishing risk-based requirements for clinical evidence. While it acknowledges the value of alternative evidence modalities—including in silico methods—it imposes more stringent demands for clinical investigations, particularly for devices incorporating novel automated, adaptive, or multi-variable closed-loop functions. This increased regulatory burden reflects the added complexity in verifying, validating, and monitoring such systems compared to traditional single-variable feedback controllers, thereby heightening the need for efficient, scalable, and regulatorily aligned methods of evidence generation.

In silico methods provide a complementary way to generate quantitative evidence by enabling controlled, repeatable, and scalable virtual studies in well-characterized

model populations. In silico clinical trials (ISCTs) can (i) extend evidence beyond narrowly defined trial cohorts, (ii) stress-test devices across broad parameter spaces and edge conditions, (iii) support early identification of potential failure modes, and (iv) focus subsequent animal and human studies on the most informative scenarios [1, 2]. For medical devices, a risk-informed credibility paradigm has emerged: ASME V&V 40 structures verification and validation around context of use and model risk [3], FDA guidance *Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions* [4] outlines expectations for clinical comparators and credibility assessment, and communities such as the VPH Institute [5, 6] and the Avicenna Alliance [6, 7] promote good simulation practice and stakeholder alignment. Mechanical ventilation is a paradigmatic application domain for such approaches. Intensive care unit (ICU) patients show pronounced heterogeneity in lung mechanics, gas exchange, respiratory drive, and comorbidities, while ventilators provide a high-dimensional control space of modes, parameters, and alarms. Clinicians must continuously integrate noisy, time-varying data to maintain gas exchange, avoid ventilator-induced lung injury and diaphragm dysfunction, and prepare for liberation from the ventilator. Despite evidence-based guidelines for core aspects such as lung-protective ventilation, adherence remains variable, and many combinations of patient characteristics and ventilator settings are insufficiently explored by conventional trials. Automation and decision-support functions in mechanical ventilation - including closed-loop control of pressure support, oxygen concentration, and other settings - aim to standardize key elements of therapy and reduce workload, but their broader adoption is limited by concerns regarding robustness across diverse pathophysiology, transparency of algorithm behaviour, and unforeseen failure modes in routine practice.

Ventilation-focused ISCTs have begun to address these challenges. Existing work spans stochastic virtual patient generators for protocol-embedded trials [8], physiology-based simulations of lung- and diaphragm-protective strategies with extracorporeal CO₂ removal [9], and digital twins in neonatal ventilation [10]. Hybrid in silico-in vitro platforms have been used to study heart-lung interactions and measurement-chain effects relevant to closed-loop testing [11], and validated cardiopulmonary models have supported systematic evaluation of closed-loop F_iO_2 control across broad physiologic and sensor conditions [12]. Model-based optimisation studies have explored mechanical power minimisation under pressure-controlled ventilation [13] and simple, interpretable strategies for patient-specific inspiratory pressure and time [14]. Collectively, these studies demonstrate the feasibility and utility of ISCTs for ventilation, but they typically focus on specific subsystems, controller concepts, or limited aspects of patient-device interaction. Moreover, most published ventilation ISCTs rely on simplified couplings between patient, ventilator, and monitoring systems that omit measurement-chain dynamics and device logic details that are critical for assessing automated functions [8–13]. To the authors knowledge, no integrated pathway has been described for end-to-end ISCTs of automated ventilation modes that coherently couples respiratory mechanics, gas exchange, respiratory control, device logic, and measurement latencies while explicitly aligning with regulatory credibility frameworks.

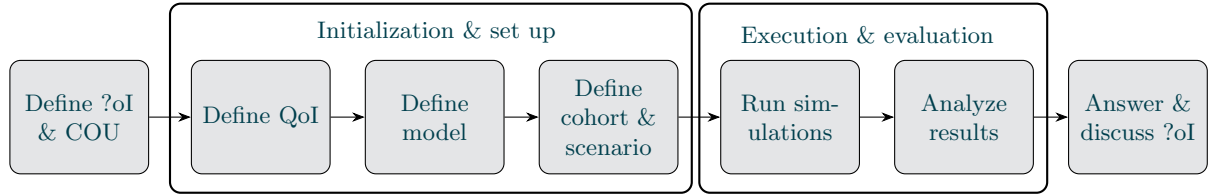


Fig. 1: Proposed workflow for an in silico clinical trial (ISCT). ?oI: Question of interest, COU: Context of use, QoI: Quantities of interest.

This work presents a patient-device model (PDM)-based ISCT framework for the pre-clinical evaluation of automated ventilator functions. It is illustrated by a case study of an automated weaning protocol (AWP) inspired by SmartCare[®]/PS (Drägerwerk AG & Co. KGaA, Lübeck, Germany [15]). Using a previously validated PDM that integrates respiratory mechanics, gas exchange, respiratory drive, and device and monitoring dynamics [16–19].

The main contributions of this work are threefold. First, a reusable, device-agnostic in silico clinical trial (ISCT) workflow aligned with ASME V&V 40 is presented for the preclinical evaluation of automated ventilator functions. Second, a methodology for virtual cohort and scenario generation is introduced that explicitly combines representative and edge-case strata to enable systematic stress-testing of controller robustness across the spectrum of mechanical ventilation. Third, the framework’s utility is demonstrated through a case study of an automated weaning protocol (AWP), including comparison with clinical comparator data.

The focus of this work lies on the framework methodology—specifically contributions 1 and 2. The case study (contribution 3) serves to validate the framework itself rather than to clinically evaluate a particular device. The AWP implemented in the ISCT is a simplified abstraction of established weaning protocols, designed to test the framework’s ability to generate transparent, decision-relevant, and regulatory-aligned evidence under real-world complexity.

The remainder of this paper is organized as follows: Section 2 describes the generic ISCT workflow and its credibility structure. Section 3 instantiates this workflow in a case study of an automated weaning protocol. Section 4 discusses the framework design, cohort methodology, and demonstrated utility. Section 5 concludes with a summary of contributions and outlook.

2 Workflow for an In Silico Clinical Trial

This section outlines the general workflow used for designing and executing in silico clinical trials. It provides a structured overview of the key components – cohort definition, parameter sampling, and patient-specific model configuration – that form the foundation for conducting reproducible and clinically meaningful simulation studies. The workflow serves as a methodological template that can be adapted to various contexts of use.

The proposed workflow of an in silico clinical trial is illustrated in Figure 1. The beginning of the process follows the structured methodology similar to the logic outlined

in the ASME V&V 40 framework [3]. The workflow can be divided into the following main steps: defining the question of interest (?oI), context of use (COU), and quantities of interest (QoI), followed by model definition, cohort and scenario definition, simulation execution, analysis of the results and the answer and discussion of the ?oI.

2.1 Generic Model Requirements

Independent of the specific application domain, the computational model employed in an in silico clinical trial (ISCT) must satisfy several generic requirements to support credible evidence generation. First, the model must exhibit physiological fidelity. It must represent the relevant physiological subsystems at a level of detail sufficient to capture the quantities of interest with acceptable accuracy. This requirement does not mandate maximal complexity. Rather, the model granularity should be commensurate with the context of use and the associated decision risk. Second, the model must support patient-specificity. It must be parameterizable to reflect inter-individual variability across the target population, with input parameters directly mapping to clinically observable or inferable patient characteristics—such as body weight, disease severity, or lung compliance. Third, for device evaluation, the model must incorporate device coupling. A faithful representation of the device under test—including its control logic, sensor dynamics, and measurement latencies—as well as the bidirectional interaction between device and patient physiology. Fourth, the model must demonstrate scenario responsiveness. It must plausibly respond to the perturbations defined in the scenario catalog (e.g. changes in mechanics, respiratory drive, or device settings), producing outputs consistent with established (patho-)physiological mechanisms. Finally, the model must ensure computational tractability. Its runtime must permit execution of the required number of simulations N within practical time and resource constraints, potentially necessitating trade-offs between model fidelity and computational cost. These requirements collectively serve as acceptance criteria during model selection or development and must be documented as part of the overall credibility assessment.

2.2 Integration of Credibility Activities

Verification and validation (V&V) activities are not confined to a single workflow step but are distributed across the entire in silico clinical trial (ISCT) lifecycle. Their placement and scope are determined by the decision risk associated with the context of use (COU), in accordance with the risk-informed credibility framework of ASME V&V 40 [3]. During model definition and development, code verification (ensuring correct implementation of the mathematical equations), calculation verification (assessing numerical accuracy and convergence), and validation—supported by comparison with clinical and literature comparator data as well as known physiological behavior—are conducted. During cohort and scenario definition, the plausibility of the generated virtual population is assessed through comparison with epidemiological data and clinical reference distributions, and scenario parameters are verified against published pathophysiological ranges. During result analysis and interpretation, the ISCT outputs are further validated against available clinical comparator data to quantify the

extent to which the trial reproduces real-world observations. Sensitivity and uncertainty analyses are also performed to characterize the influence of model assumptions and input uncertainties on the quantities of interest. The cumulative credibility evidence derived from these activities determines whether the model is fit for its intended purpose within the defined COU. A detailed credibility assessment for the specific patient-device model used in this case study is provided in Hennigs et al. [16].

2.3 Workflow steps

I. Question of Interest

The starting point of an in silico clinical trial is the definition of the question of interest. This question specifies what is to be investigated within the in silico clinical trial. Typical examples include the assessment of device performance under physiological variability, comparison of alternative therapeutic strategies, or the evaluation of patient-specific responses. In the context of the ISCT, the ?oI is defined at the level of the device or function under test and determines the overall trial objective. Multiple questions in an ISCT are possible.

II. Context of Use

Based on the ?oI, the context of use is defined. The context of use describes how the model will be used to address the question of interest, specifying the clinical or physiological setting, the type of analysis, and the intended role of the model in the decision-making process. A clearly defined COU is essential, as it directly influences the model requirements, the acceptable level of uncertainty, and the design of the virtual experiment.

III. Quantities of Interest

From the ?oI and COU, the relevant quantities of interest are derived. These represent the model outputs that will be analyzed to answer the question of interest. Typical examples include hemodynamic variables such as arterial blood pressure or heart rate, ventilatory variables such as tidal volume or blood gas analysis parameter, or performance metrics of the device under test.

IV. Model Definition and Development

Once the ?oI, COU, and QoI are defined, the model design can be specified. This includes identifying the required physiological subsystems, relevant input parameters, and interconnections. The model must satisfy the generic requirements outlined in Section 2.1, with the level of fidelity and complexity determined by the COU and decision risk. Depending on the available model base, this step may involve model development, implementation, and verification and validation (V&V) activities as described in [16, 20].

V. Cohort and Scenario Definition

Building on both the defined ?oI and COU, the virtual cohort is established to represent the target patient population relevant to the clinical question under investigation. The ?oI determines which patient characteristics and pathophysiological conditions must be captured, while the COU specifies the clinical setting, constrains the intended scope, and defines boundary conditions for cohort design. Cohort definition includes the selection of patient characteristics such as age,

body size, weight, disease state, or other relevant physiological features. In parallel, the simulation scenarios are defined to describe how the cohort will be exposed to different conditions or interventions over time – such as varying ventilatory settings, drug dosages, or disease progressions – directly reflecting the conditions and risk cases identified in the ?oI and COU.

VI. Parameter Sampling and Simulation Execution

For each cohort member, the model input parameters are sampled according to predefined probability distributions. The sampling strategy (e.g. random sampling, Latin hypercube sampling, or other design of experiment methods) is chosen to sufficiently cover the physiological variability defined by the cohort and scenario. The number of simulations N is a critical design decision that must be explicitly justified with respect to the ?oI: for exploratory or sensitivity analyses, N may be determined by convergence of key output statistics across incrementally larger samples. For hypothesis-testing objectives, standard statistical power calculations based on the quantities of interest and the expected effect size should be applied and for coverage-oriented designs such as Latin hypercube sampling, N should ensure adequate space-filling across the relevant input parameter ranges. Computational budget and available model runtime impose additional practical constraints. Once all model instances are configured, the simulations are executed and the results are stored for subsequent analysis.

VII. Result Analysis

The final step involves the statistical and physiological analysis of the simulation results in relation to the ?oI. This includes descriptive statistics, uncertainty quantification, and sensitivity analysis, depending on the trial objective. The outcome of this step provides insight into the expected range of system or device behavior across the virtual population and supports quantitative conclusions about performance, safety, or efficacy.

VIII. Answer and Discuss the Questions of Interest

In the final step, the analysis results are synthesized to explicitly answer the originally defined questions of interest. This includes interpreting the simulation outcomes in the context of the COU, assessing whether the ?oI meet predefined acceptance criteria across the virtual cohort, and drawing conclusions regarding device performance, safety, or efficacy. Limitations arising from model assumptions, cohort representativeness, and scenario coverage must be transparently discussed. If the results do not sufficiently address the ?oI, iterative refinement of the trial design may be indicated. This step provides the basis for regulatory or engineering decision-making and identifies remaining gaps requiring further investigation.

The sequential ordering of the workflow reflects a deliberate design principle: each step constrains and informs all subsequent steps, ensuring that downstream decisions remain traceable to the original clinical question. The ?oI determines *what* must be answered. The COU specifies *how* and *under which conditions* the model will be used. The QoI translate these requirements into measurable model outputs. Only after these three elements are fixed the model complexity, cohort composition, and scenario

catalog be meaningfully can be defined – avoiding the common pitfall of building a model first and then searching for questions it might answer. This top-down logic mirrors the risk-informed approach of ASME V&V 40 [3], where the required level of model credibility is derived from the decision risk associated with the COU rather than from the model itself.

3 In Silico Clinical Trial of an Automated Weaning Protocol

This section instantiates the generic ISCT workflow (Section 2) through a case study of an automated weaning protocol for ICU ventilation. Table 1 maps each workflow step to its concrete realization in the present case study, providing traceability between the generic methodology and the specific implementation.

Table 1: Mapping of generic workflow steps (Section 2) to case study realization.

Workflow step	Case study realization	Section
Define ?oI & COU	5 questions on AWP performance, lung protection, asynchrony, oscillations, and influence factors	3.1
Define QoI	V_T , RR, etCO ₂ , P_{plat} , asynchrony events, P_{insp} oscillations	3.1
Define model	Previously validated PDM [16, 19] with simplified AWP	3.2.1
Define cohort & scenario	400 virtual ICU patients, 9 perturbation scenarios	3.2.2, 3.2.3, 3.2.4
Run simulations	3600 runs (400 patients $\times$ 9 scenarios), LHS + uniform edge-case sampling	3.3
Analyze results	TiT, overshoot, LET, composite score, Spearman correlation	3.4
Answer & discuss ?oI	Summary of findings per ?oI, comparison with clinical data	3.6

As a representative use case, the AWP is inspired by SmartCare[®]/PS (Drägerwerk AG & Co. KGaA, Lübeck, Germany) [15], a clinically established and commercially available pressure-support weaning system. This protocol was selected for three reasons: (i) it is one of the most thoroughly documented AWP in the clinical literature, providing a well-defined algorithmic basis for abstraction into the simulation framework, (ii) it is widely used in clinical practice and therefore represents a practically relevant and regulatory-pertinent evaluation target, and (iii) clinical ventilation data recorded during SmartCare[®]/PS weaning are available [21], enabling a direct quantitative comparison between in silico results and real-world observations.

Using the validated patient-device model [16, 19], a virtual cohort is constructed, scenario-specific parameters are defined, and a full simulation campaign is conducted.

The case study demonstrates the practical implementation of the workflow and highlights its utility for assessing patient-device interactions and algorithmic performance within a defined ?oI and COU.

3.1 Question of Interest, Context of Use and Quantities of Interest

3.1.1 Question of Interest

For the use case of evaluating the AWP in an in silico clinical trial, the following questions of interest are formulated:

- I. Does the automated weaning protocol maintain tidal volume V_T , respiratory rate RR, and end-tidal carbon dioxide $etCO_2$ within the predefined acceptable working ranges over time?
- II. Do the resulting tidal volume V_T and plateau pressure P_{plat} satisfy guideline-based criteria for lung-protective ventilation?
- III. Does patient-ventilator asynchrony occur during operation with the automated weaning protocol? If so, what types and how often?
- IV. Do inspiratory pressure oscillations occur during operation with the automated weaning protocol? If so, how often?
- V. What are the dominant patient and ventilator parameters influencing the performance of the automated weaning protocol (AWP), and how do they affect key outcomes such as tidal volume control, patient-ventilator synchrony, and lung-protective ventilation?

3.1.2 Context of Use

The aim of the ISCT is to evaluate in silico whether the automated weaning protocol can automatically implement a clinical therapeutic strategy under predefined conditions and to assess its impact on ventilatory performance and patient-ventilator interaction. In particular, the ISCT is used to quantify how well the AWP maintains tidal volume, respiratory rate, and end-tidal CO_2 within the predefined acceptable ranges, whether the resulting combinations of V_T and plateau pressure P_{plat} are consistent with guideline-based lung-protective ventilation, and how often patient-ventilator asynchrony and P_{insp} oscillations occur across different scenarios. It enables investigation of device behavior in a physiologically and pathophysiologically validated environment prior to clinical testing.

This COU describes a preclinical modeling and simulation workflow intended to support two complementary regulatory objectives: (i) preclinical design evaluation and iterative refinement of AWP prior to first-in-human testing, serving as a computational alternative or complement to bench and animal studies for assessing controller robustness, edge-case behavior, and potential failure modes across heterogeneous patient populations; and (ii) post-market clinical follow-up (PMCF) under EU MDR 2017/745, where in silico evidence can augment real-world performance data by systematically probing operating conditions, patient subgroups, or rare adverse scenarios that are under-represented in routine clinical surveillance. The framework is not

intended to replace clinical investigations and does not demonstrate clinical effectiveness or safety in a regulatory sense, nor should simulation results be used to guide individual patient care. All model assumptions, simplifications, uncertainties, and boundary conditions must be explicitly documented and justified within the associated credibility plan in accordance with ASME V&V 40 and applicable FDA and EU guidance.

3.1.3 Quantities of Interest

Tidal volume V_T , respiratory rate RR, and end-tidal CO_2 etCO_2 , plateau pressure P_{plat} , and quantitative descriptors of patient-ventilator asynchrony, including the incidence (number of events) and relative frequency of specific asynchrony types and quantitative descriptors of P_{insp} oscillations.

3.2 Model Definition, Virtual Cohort and Scenario Generation

This section details the design of the patient-device model (PDM), the construction of a heterogeneous virtual patient cohort, and the formulation of clinically relevant scenarios for the preclinical evaluation of an automated weaning protocol (AWP). The overarching objective is to systematically assess the robustness, safety, and performance of the AWP under a broad spectrum of physiological conditions spanning typical ICU presentations as well as extreme, high-risk scenarios. Thereby enabling a credibility-aware in silico clinical trial (ISCT) aligned with ASME V&V 40 [3].

3.2.1 Model Definition and Development

The patient-device model utilized in this study and its requirements for the described COU has been summarized in detail in published papers [17–19]. It integrates respiratory mechanics, gas exchange, and respiratory control, along with device and monitoring dynamics.

A simplified automated weaning protocol was abstracted from [15] as an automated P_{insp} controller that adapts support levels to meet ranges for V_T , RR and etCO_2 in CPAP/CSV-PS mode. Device-specific flow triggering β and off-cycling sensitivity α are varied within clinically realistic ranges as well as the initial value for inspiratory pressure P_{insp} , PEEP and inspiratory O_2 fraction $F_{\text{I}}\text{O}_2$ as well as the patient sedation level scaling factor ρ_{sed} (Table 2).

3.2.2 Virtual Cohort

The virtual cohort was designed to represent typical adult ICU patients [22–25] who are eligible for assisted mechanical ventilation, while also covering clinically relevant edge cases.

Demographics and Anthropometry

The cohort included representative adult ICU patients with diverse age, sex, and anthropometric characteristics (Table 3). The median age is 75 years [22]. The age-dependence of ICU admission rates was modeled using a Gamma distribution based

Table 2: Initial ventilator parameters and sedation level at baseline (prior to automated adaptation) for representative and edge cases. P_{insp} : Inspiratory pressure / driving pressure, PEEP: Positive end-expiratory pressure, $F_i\text{O}_2$: Fraction of inspiratory O_2 , α : Off-cycling sensitivity, β : Flow trigger, ρ_{sed} : Sedation level scaling factor.

Parameter	Representative	Edge case
P_{insp} [mmHg]	8 – 14	5 – 20
PEEP [mmHg]	5 – 10	3 – 12
$F_i\text{O}_2$ [%]	30 – 50	21 – 60
α [%]	25 – 35	20 – 50
β [L/min]	1 – 1.5	0.5 – 3
ρ_{sed}	0 – 0.4	0.6 – 0.8

Table 3: Demographics and anthropometry (representative strata). Continuous variables (age and height) are given as mean $\pm$ standard deviation, while categorical variables (sex and BMI) are represented as percentage distributions across strata.

Feature	Population characteristics	Literature
Age [years]	75.4 ± 15.7	[22]
Sex	64 % male / 36 % female	[22, 24]
Height [m]	1.76 ± 0.07 male, 1.64 ± 0.07 female	[26]
Body Mass Index (BMI) [kg/m^2]	<20 (5 %), 20 – 25 (40 %), 25 – 30 (35 %), 30 – 39 (15 %), ≥ 40 (5 %)	[25]

on the study by Garland et al. [22]. The Gamma probability density function is

$$f(x; k, \theta) = \frac{1}{\Gamma(k) \theta^k} x^{k-1} e^{-x/\theta}, \quad x > 0, \quad (1)$$

where the parameters $k = 11.69$ and $\theta = 5.88$ were estimated. The sex distribution was typical of ICU populations (64 % male / 36 % female) [23, 24]. Heights and weights were sampled from truncated normal distributions. Obesity was explicitly included as an edge factor, with 20 % of patients having a body mass index (BMI) $\geq 30 \text{ kg}/\text{m}^2$ [25].

Clinical and Pathophysiological Strata

The virtual cohort incorporated representative pathophysiological conditions commonly encountered in mechanically ventilated patients, along with edge cases relevant for stress-testing automated ventilator control (Table 4). Strata includes acute respiratory distress syndrome (ARDS) (30 %), chronic obstructive pulmonary disease (COPD) (30 %), postoperative respiratory compromise (25 %), neuromuscular weakness (10 %), and congestive heart failure with pulmonary edema (5 %).

Table 4: Clinical and pathophysiological strata of ICU patients in the virtual cohort with parameter adaptation and edge cases. For acute respiratory distress syndrome (ARDS) and chronic obstructive pulmonary disease (COPD) parameter choice see Tables A1 and A2, respectively. ρ_{sed} : Sedation level scaling factor, ρ_{resp} : Respiratory failure scaling factor, f_s : Shunt fraction, $E_{\text{max,lv}}$ and $E_{\text{max,rv}}$: Maximum elastance of the left and right ventricle, respectively.

Stratum	Share	Representative	Edge cases
ARDS	30 %	mild, moderate	severe
COPD	30 %	GOLD I–III	GOLD IV
Postoperative	25 %	$\rho_{\text{sed}} = 0.6$	$\rho_{\text{sed}} = 0.8$
Neuromuscular	10 %	$\rho_{\text{resp}} = 0.4$	$\rho_{\text{resp}} = 0.6$
Lung edema, heart failure	5 %	$f_s = 0.25$, $E_{\text{max,lv}} = 3.5$ mmHg/mL, $E_{\text{max,rv}} = 2.2$ mmHg/mL	$f_s = 0.35$, $E_{\text{max,lv}} = 4$ mmHg/mL, $E_{\text{max,rv}} = 2.7$ mmHg/mL

The clinical and pathophysiological strata, together with the corresponding parameter adaptations and edge-case definitions, are summarized in Table 4. Sedation depth is modeled via the scaling factor ρ_{sed} , which is varied in the postoperative stratum (representative: $\rho_{\text{sed}} = 0.6$, edge case: $\rho_{\text{sed}} = 0.8$) to reflect different levels of sedation, where $\rho_{\text{sed}} = 0$ denotes a fully awake state and $\rho_{\text{sed}} = 1$ a fully sedated state. Respiratory muscle weakness is captured by the respiratory failure scaling factor ρ_{resp} in the neuromuscular stratum (representative: $\rho_{\text{resp}} = 0.4$, edge case: $\rho_{\text{resp}} = 0.6$). In the lung-edema/heart-failure stratum, shunt fraction f_s and ventricular elastances $E_{\text{max,lv}}$ and $E_{\text{max,rv}}$ are adapted to represent varying severity of cardiopulmonary impairment. Particularly for the selected edge cases, it is important to note that such patients (e.g. with severe ARDS or heart failure) would typically not be considered ready for weaning in routine clinical practice. However, the intended use of SmartCare®/PS [15] does not formally exclude such cases a priori. Therefore, these edge cases were deliberately included in the virtual cohort to explore the full range of clinically plausible pathophysiological conditions and to assess controller behavior also in scenarios where automated weaning would be undesirable or potentially unsafe. This allows the in silico clinical trial to cover the entire relevant parameter space and to identify conditions under which additional safety constraints or exclusion criteria may be required.

3.2.3 (Patho-)Physiological Parameter Selection

Based on the generated virtual cohort, the corresponding model input parameters for lung mechanics and gas exchange submodels are determined according to each patient’s specific characteristics. Model input parameters that depend on individual features - such as age, height, sex, weight, and relevant pathological conditions - were derived from established physiological relationships or empirical formulations reported in the literature. Where applicable, standard deviations were defined to represent inter-individual variability and to enable stochastic sampling around the mean parameter values.

Table A3 provides an overview of the patient-specific correlations used for selected

Table 5: Scenario catalog and perturbations with model parameter changes. E_L : Lung elastance, R : Airway resistance, ρ_{resp} : Respiratory failure factor, f_s : Shunt fraction, V_D : Anatomical dead space, f_{V_D} : Pulmonary dead space, ρ_{sed} : Sedation level scaling factor.

ID	Scenario	Perturbation	Parameter
S1	Basic	No event	–
S2	Lung mechanics I	Increase in lung elastance E_L	$E_L = E_{L,\text{calc}} \cdot 2$
S3	Lung mechanics II	Increase in airway resistance R	$R = R_{\text{calc}} \cdot 2$
S4	Drive shift I	Increase in respiratory failure ρ_{resp}	$\rho_{\text{resp}} = 0.5$
S5	Drive shift II	Increase in sedation level ρ_{sed}	$\rho_{\text{sed}} = 0.6$
S6	Gas exchange I	Increase in shunt fraction f_s	$f_s = 0.4$
S7	Gas exchange II	Increase in anatomical dead space V_D	$V_D = 0.3 \text{ L}$
S8	Gas exchange III	Increase in pulmonary dead space f_{V_D}	$f_{V_D} = 0.4$
S9	Weaning	Sedation taper	$\rho_{\text{sed}} = [0.4, 0.2, 0]$

model parameters, including their mathematical expressions and respective literature sources. Parameters related to disease-specific alterations in lung mechanics and gas exchange are summarized separately in Tables A1 and A2. The severity of COPD was classified according to the established GOLD criteria [27], whereas for ARDS, representative parameter sets were defined for mild, moderate, and severe conditions based on the Berlin Definition [28].

3.2.4 Scenario Generation

To evaluate the model’s robustness, stability, and synchrony with respect to the four predefined ?oI, several exemplary test scenarios were developed, as summarized in Table 5. Each scenario specifies a distinct set of parameter perturbations targeting the pathophysiological dimensions identified in the ?oI: changes in lung mechanics (S2, S3), respiratory drive (S4, S5), and gas exchange (S6–S8), as well as a clinically representative sedation-taper weaning sequence (S9). In all cases, all quantities of interest are evaluated.

The total simulation duration was set to 2000 s, with perturbations introduced at $t = 800 \text{ s}$, leaving 1200 s of post-perturbation observation time. Given that the AWP adjusts P_{insp} every 120–300 s, this window accommodates four to ten complete controller adjustment cycles, which is sufficient to assess transient controller response, convergence to a new steady state, and the emergence of oscillatory behavior as defined in the ?oI. For weaning scenario S9, perturbation steps are introduced at 500 s, 1000 s, and 1500 s, providing equal 500 s observation intervals per sedation level – each covering at least four controller cycles – to capture the progressive response to sedation taper. While a longer simulation horizon would increase absolute excursion duration and allow for rarer events, the 2000 s window was deemed a pragmatic and sufficient choice for preclinical controller characterization within this COU. Longer scenarios representing multi-hour ICU stays are identified as future work.

3.3 Cohort Sampling and Simulation Execution

A hybrid sampling strategy was implemented to capture both typical and edge-case physiological conditions within the virtual patient cohort. Representative patient parameters - including demographics, lung mechanics, gas exchange, and device settings - were sampled using latin hypercube sampling (LHS) to efficiently span the central ranges observed in ICU populations. To ensure the inclusion of rare but safety-critical conditions, edge cases were additionally sampled using uniform approaches [12]. The virtual cohort comprised 400 simulated patients, with 60 % representing typical cases and 40 % covering edge cases. Each virtual patient was simulated under nine predefined scenarios (S1 – S9) defined in Table 5, resulting in 3600 total simulations.

3.4 Result Analysis

Quantities of interest (V_T , RR, etCO_2 , P_{plat} , asynchrony, and P_{insp} oscillation) are evaluated for each simulation run and scenario to quantify ventilatory performance and patient-ventilator interaction.

3.4.1 Evaluation Metrics

To quantify ventilatory performance and patient-ventilator interaction, the following metrics were computed for each simulation run.

Time in target (TiT) is defined as the fraction of simulation time during which a quantity of interest remains within its predefined target range. Overshoot quantifies the maximum exceedance of the upper target limit, expressed as a percentage of that limit. The longest excursion time (LET) represents the maximum continuous duration for which the QoI remains outside the target range. The composite performance score is a normalized, equally weighted average of TiT, overshoot, and LET across all QoI, yielding a scalar in $[0, 1]$ for each scenario.

Asynchrony detection classifies breaths into synchronous triggers, delayed triggers, ineffective triggers, auto-triggers, and double efforts based on the methodology by Sauer et al. [29]. Inspiratory pressure oscillations are identified via k -means segmentation of the inspiratory pressure time series P_{insp} . Simulations with more than three valid oscillations—defined as transitions between low and high pressure states lasting between 120 and 300 seconds, matching the controller’s adjustment interval—are classified as oscillatory. Influence analysis employs Spearman rank correlation to assess the association between all patient input variables and outcome metrics.

Target ranges were defined as follows: tidal volume $V_T > 300 \text{ mL}$, respiratory rate RR: $15\text{--}30 \text{ min}^{-1}$, end-tidal carbon dioxide $\text{etCO}_2 < 55 \text{ mmHg}$ (based on [15]), and lung-protective criteria: $V_{T,\text{LP}}: 6\text{--}8 \text{ mL/kg}$ ideal body weight (IBW), $P_{\text{plat},\text{LP}} \leq 30 \text{ cmH}_2\text{O}$ [30]. Formal metric definitions are provided in Appendix A.2.

3.4.2 Clinical Data Analysis for Comparison

As part of the credibility assessment, simulation results were compared with retrospective clinical ventilation data from [21]. The clinical data used in this work originated from a prospective, single-centre observational study conducted in three surgical intensive care units at University Hospital Schleswig-Holstein (UKSH), Campus Kiel, Kiel,

Germany. The study protocol, including the collection and secondary use of the clinical data for the modelling analyses presented in this study, was approved by the Ethics Committee of the Medical Faculty of Christian-Albrechts-Universität zu Kiel, Germany (approval no. D411/13). This committee is the responsible ethics-review body for research conducted at UKSH Campus Kiel under the cooperation model between the Medical Faculty of Christian-Albrechts-Universität zu Kiel and UKSH. All methods involving human participants and human data were carried out in accordance with relevant institutional, national, and legal guidelines and regulations. The requirement for informed consent was waived by the Ethics Committee. All study data were analysed anonymously. Clinical recordings were obtained from mechanically ventilated patients during weaning with the Dräger SmartCare®/PS protocol and analyzed using the same outcome metrics as applied in the in silico experiments.

The dataset consisted of multiple measurement files per patient, with recordings differing in duration and availability of physiological variables. All files were screened for data integrity and only recordings containing valid time-resolved measurements were included in the analysis. For each recording V_T , RR, $etCO_2$, P_{insp} and P_{plat} were extracted when available.

3.5 Results of the Case Study

3.5.1 Generated Cohort

The virtual patient cohort ($N = 400$) captured a broad spectrum of ICU-relevant physiological and demographic characteristics (Figure 2a). The age distribution was centered around a median of 75 years, spanning 40 – 100 years. Height and BMI followed the expected sex-dependent distributions, with BMI classes ranging from underweight to severe obesity. Sex distribution reflected the target population.

The disease sampling (Figure 2a) ensured both representative and edge-case patients. ARDS, COPD, postoperative, neuromuscular, and lung edema were included according to the predefined proportions. The stacked barplot of disease severity confirmed the presence of both common and extreme parameter values, illustrating the cohort’s heterogeneity.

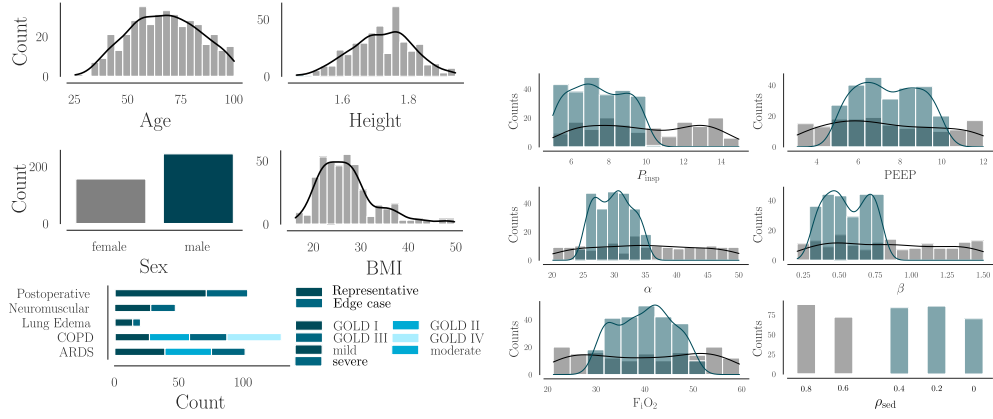
Initial ventilation parameters were sampled to reflect typical ICU settings and extreme conditions. Figure 2b illustrates these distributions.

3.5.2 Performance Metrics, Composite Score, Asynchrony, and Inspiratory Pressure Oscillations

Performance metrics across all nine scenarios are summarized in Figures 3, 4a, and 4b, and detailed in Appendix A.3.1.

Time in target (TiT) for tidal volume ranged from 49.4 % (S2) to 84.5 % (S7), while respiratory rate (RR) remained stable with TiT exceeding 84 % across all scenarios. End-tidal carbon dioxide was well maintained except in scenario S6 (shunt) and S7 (dead space), where TiT dropped to 33.0 % and 54.7 %, respectively.

For lung-protective ventilation, $V_{T,LP}$ TiT ranged from 41.0 % to 59.2 %. Plateau pressure for lung-protective ventilation ($P_{plat,LP}$) achieved near 100 % TiT in scenarios S1 and S2 but decreased to 44.0 % in S6, indicating substantial difficulty in maintaining



(a) Characteristics of the virtual patient cohort ($N = 400$). Histograms with kernel density estimates for age [years], height [m], and BMI [kg/m²], illustrating the distributions of continuous patient parameters and categorical distributions: Sex (male/female) and disease severity (stacked barplot). Edge-case patients are included.

(b) Distribution of ventilation parameters ($N = 400$) across the virtual patient cohort. Representative patients (60%, black) and edge-case patients (40%, blue) are shown. Ventilation parameters (inspiratory pressure P_{insp} [cmH₂O], PEEP [cmH₂O], inspiratory fraction of O₂ $F_{\text{I}O_2}$ [%], off-cycling sensitivity α [%], and flow-trigger β [L/min]) and sedation level ρ_{sed} are displayed as histograms with kernel density estimates.

Fig. 2: Characteristics of the virtual patient cohort and corresponding ventilation parameter distributions used for the in silico clinical trial.

lung protection under severe gas exchange impairment.

The composite performance score was highest for scenario S9 (weaning) at 0.81 and lowest for scenario S6 (shunt) at 0.44, reflecting pronounced imbalances between individual metrics under extreme pathophysiological conditions.

Asynchrony analysis revealed delayed triggers as the most prevalent type across scenarios, while ineffective triggers dominated in S2 (stiff lung). Overall, 11.1 % of simulations exhibited inspiratory pressure (P_{insp}) oscillations, which were primarily associated with ARDS and correlated with disease state (Spearman $\rho = 0.25$). This frequency is comparable to clinical observations (15.5 %), supporting the framework's ability to reproduce emergent controller behavior.

3.5.3 Influence Analysis

Spearman correlation analysis (Figure A1) identified disease state, sedation level, and initial inspiratory pressure as the dominant determinants of AWP performance. ARDS was associated with impaired lung-protective control ($\rho = -0.34$ for $P_{\text{plat,LP}}$ TiT), while deeper sedation reduced asynchronous events ($\rho = -0.41$ for double effort). Demographic variables (age, height, BMI) showed weak associations ($|\rho| < 0.27$). A detailed breakdown is provided in Appendix A.3.2.



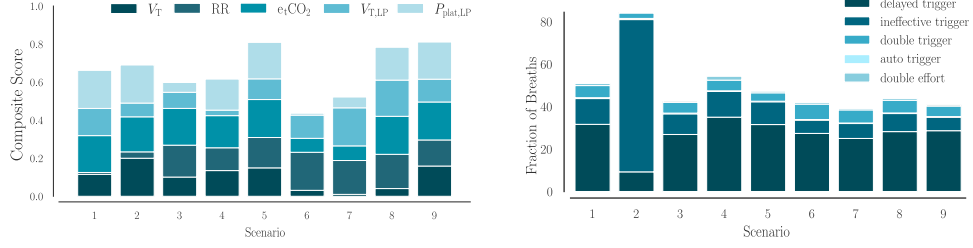
Fig. 3: Performance heatmaps for the acceptable working ranges and lung protective ventilation scenarios. Top: Acceptable working ranges of tidal volume V_T , respiratory rate RR, and end-tidal carbon dioxide etCO₂. Bottom: Lung protective metrics of tidal volume for lung protective ventilation $V_{T,LP}$ and plateau pressure for lung protective ventilation $P_{plat,LP}$. Each heatmap shows three metrics: Median over all patients of time in target [%], overshoot [%], and longest excursion (% of total simulation time, 2000 s).

These findings provide a basis for subsequent, more targeted analyses by refining the virtual cohort through denser sampling in identified critical parameter regions.

3.6 Interpretation of Results in Relation to the Questions of Interest

Table 6 synthesizes the simulation results with respect to the five predefined questions of interest. The AWP generally maintains acceptable working ranges for the primary QoI, with transient deviations that resolve within a few controller cycles. However, lung-protective ventilation remains challenging in scenarios with severely impaired gas exchange or respiratory mechanics, revealing a fundamental trade-off between adequate CO₂ clearance and pressure limitation. This trade-off – identified systematically through the scenario catalog – would require multiple independent clinical studies to characterize with conventional methods.

Patient-ventilator asynchrony occurs in a phenotype- and scenario-dependent manner, with rates comparable to clinical literature [29, 31]. The observed P_{insp} oscillation rate (11.1 %) aligns with clinical data (15.5 %), demonstrating the framework’s ability to



(a) Results of the in silico clinical trial for composite scores. The stacked bar plot displays performance metrics (time in target, overshoot, and longest excursion) across ventilatory parameters: tidal volume V_T , respiratory rate RR , and end-tidal CO_2 e_tCO_2 , tidal volume for lung protective ventilation $V_{T,LP}$ and plateau pressure for lung protective ventilation $P_{plat,LP}$.

(b) Results of the in silico clinical trial for asynchrony profile of patient-ventilator interaction. The stacked bar plot shows the fraction of breaths [%] classified as different types of asynchrony - delayed trigger, ineffective trigger, double trigger, auto-trigger, and double effort. Breaths with normal synchrony are not included in the plot.

Fig. 4: In silico clinical trial results for ventilatory performance and patient-ventilator interaction for each scenario. (a) Composite performance scores for selected ventilatory targets. (b) Distribution of patient-ventilator asynchrony types across simulated breaths.

reproduce emergent controller behavior. A detailed interpretation per ?oI is provided in Appendix A.3.3.

3.7 Comparison with Clinical Dataset

The in silico cohort reproduces key ventilatory patterns observed in clinical data [21]. Demographics are well aligned (Table 7). Simulated TiT values are generally higher and variability lower than in clinical recordings, which is expected given the absence of real-world confounders (leakage, pain, medication changes, sleep-state transitions). Importantly, the qualitative trends – including the relative difficulty of lung-protective V_T control and the occurrence of P_{insp} oscillations – are consistent between in silico and clinical observations.

4 Discussion

The results presented in Section 3.5 collectively demonstrate the framework’s capacity to generate quantitative, clinically plausible results across all five ?oIs. The following discussion interprets these findings not at the level of individual simulation outcomes, but at the level of the framework methodology: what do these results reveal about the design choices made in Sections 2 and 3.2, and what do they imply for the reusability and regulatory applicability of the proposed ISCT workflow?

Table 6: Summary of ISCT findings per question of interest.

#	Question of Interest	Key Finding	Framework Implication
1	Acceptable working ranges maintained?	V_T , RR, etCO_2 generally within target; deviations transient (<80 s); exceptions in S6/S7 (gas exchange impairment)	Scenario catalog successfully identifies conditions where AWP reaches its limits
2	Lung-protective criteria met?	$V_{T,LP}$ near upper limit; P_{plat} exceeded in S3/S6/S7; fundamental tension between gas exchange and lung protection	Edge-case strata essential for detecting safety-relevant trade-offs
3	Patient-ventilator asynchrony?	Delayed triggers 20–35 %; ineffective triggers 5–10 % (up to 80 % in S2); scenario- and phenotype-dependent	Cohort heterogeneity enables identification of asynchrony-prone subpopulations
4	P_{insp} oscillations?	11.1 % of simulations oscillatory (clinical: 15.5 %); ARDS-associated	ISCT reproduces emergent controller behavior without clinical risk
5	Dominant influencing parameters?	Disease state, sedation, initial P_{insp} dominant; demographics secondary	Influence analysis guides targeted cohort refinement for subsequent ISCTs

4.1 Framework Design and Alignment with ASME V&V 40

The proposed ISCT workflow makes a deliberate methodological choice that distinguishes it from the majority of existing ventilation ISCTs: the top-down sequencing from ?oI to COU to QoI to model constrains all downstream decisions and ensures that the entire trial design remains traceable to an explicit clinical or regulatory question. This top-down logic mirrors the risk-informed structure of ASME V&V 40 [3], where required model credibility is derived from decision risk rather than from model complexity alone. In contrast, most published ventilation ISCTs are bottom-up: a validated model is built first, and questions are subsequently identified that the model might answer [8–13]. While this is a pragmatically understandable approach, it introduces a risk of confirmation bias and limits the generalizability of individual simulation results to other contexts of use.

A second distinguishing feature is the explicit separation between generic framework components (Section 2) and case-specific instantiations (Section 3). This separation is not merely presentational: it enables the framework to function as a reusable template. Any automated ventilation strategy – including closed-loop $F_i\text{O}_2$ control [12], adaptive PEEP optimization, or AI-driven multi-variable controllers as envisioned by IEC 60601-1-10 – can be evaluated by re-instantiating the same workflow with a different ?oI, a different COU, and a different device model, while the cohort generation methodology, scenario catalog design principles, and credibility activities remain unchanged.

Table 7: Demographics of clinical comparator data [21] and results of the performances metrics in comparison with in silico clinical trial results (mean and standard deviation).

Feature	Clinical data	ISCT data
Age [years]	70.3 ± 11.6	75.4 ± 15.7
Sex	68 % male / 32 % female	64 % male / 36 % female
Height [m]	1.73 ± 0.10	1.73 ± 0.09
Body Mass Index (BMI) [kg/m ²]	26.0 ± 5.2	25.9 ± 6.6
V_T TiT [%]	94.32 ± 4.61	69.74 ± 11.30
V_T Overshoot [%]	0 ± 0	0 ± 0
V_T Longest Excursion [s]	*	108.0 ± 35.60
RR TiT [%]	97.9 ± 21.2	95.7 ± 4.1
RR Overshoot [%]	6.7 ± 26.4	17.9 ± 4.5
RR Longest Excursion [s]	*	17.3 ± 4.8
etCO ₂ TiT [%]	81.8 ± 40.4	93.2 ± 19.7
etCO ₂ Overshoot [%]	0 ± 10.6	0 ± 3.9
etCO ₂ Longest Excursion [s]	*	59 ± 265
$V_{T,LP}$ TiT [%]	38.54 ± 23.28	51.3 ± 6.2
$V_{T,LP}$ Overshoot [%]	93.2 ± 73.3	30.0 ± 9.2
$V_{T,LP}$ Longest Excursion [s]	*	135 ± 40.9
$P_{plat,LP}$ TiT [%]	91.22 ± 25.36	83.32 ± 21.80
$P_{plat,LP}$ Overshoot [%]	9.47 ± 23.11	6.12 ± 6.67
$P_{plat,LP}$ Longest Excursion [s]	*	176.6 ± 241.4
P_{insp} Oscillations [%]	15.5	11.1

* Not directly comparable: clinical recording durations (hours to days) substantially exceed the simulation horizon of 2000 s, rendering absolute longest-excursion values incommensurable.

This modularity is a precondition for regulatory scalability: it allows submission of a single framework description with device-specific instantiation documents, rather than requiring a full methodological justification for each new device evaluation.

The integration of credibility activities distributed across all workflow steps – rather than confined to a single validation phase – is directly aligned with ASME V&V 40’s principle that model credibility is a continuous property, not a binary certification. Code verification, cohort plausibility checks, and comparison with clinical comparator data are each positioned at the workflow step where they are most informative, ensuring that credibility evidence accumulates throughout the trial rather than only at its conclusion.

4.2 Virtual Cohort and Scenario Design as Framework Elements

The virtual cohort design encodes two methodological principles that are framework-level contributions rather than case-study-specific choices.

First, the representative–edge-case stratification (60%/40%) reflects a deliberate two-objective design: representative cases provide the basis for characterizing expected

device performance across the target population, while edge cases probe the boundaries of controller robustness in conditions that are clinically plausible but rarely encountered in conventional trials. This dual objective cannot be satisfied by a single sampling strategy. It requires an explicit stratum definition that ties clinical risk categories (here: severe ARDS, GOLD IV COPD, deep sedation) to specific parameter regions and ensures their inclusion regardless of their prevalence in the overall cohort. This principle is transferable to any device evaluation: the specific strata will differ by domain, but the design logic – define representative and edge-case strata explicitly, and sample them separately – is a generic recommendation of the framework. Second, the hybrid LHS/uniform sampling strategy serves a distinct purpose from either pure random sampling or full factorial design. LHS ensures efficient coverage of the central physiological space with a moderate number of simulations, while uniform edge-case sampling guarantees that the boundaries of the parameter space are populated independently of probability mass. For exploratory ISCTs at the scale of $N = 400$, this combination is methodologically more efficient than increasing N with pure random sampling, which would predominantly generate additional central-range patients. The cohort size of $N = 400$ was selected to enable a proof-of-concept demonstration of the complete framework, including influence analysis and clinical comparator validation, within a practical computational budget. It is not a definitive sample size for a statistically powered study. Rather, it is a first-phase screening cohort in a phased ISCT strategy in which subsequent iterations will employ larger, adaptively sampled cohorts in the critical parameter regions identified here.

The decision to exclude correlated parameters and multi-morbidity in the current design was deliberate: it ensures clean attribution of performance differences to individual pathophysiological conditions and preserves interpretability of the influence analysis. Future extensions of the framework will incorporate correlated parameter sampling and multi-morbidity to increase clinical realism, at the cost of reduced interpretability.

4.3 Framework Utility Demonstrated Through the Case Study

The case study of the automated weaning protocol serves as a proof of concept for the framework’s ability to generate decision-relevant, regulatory-aligned evidence under conditions of real-world complexity. The following framework properties were directly tested and confirmed through the case study instantiation.

Table 1 maps each of the five questions of interest to a specific workflow step, analysis method, and result section. The full chain from clinical question to quantitative answer is thus explicitly documented and reproducible. This end-to-end traceability is a prerequisite for regulatory use: it enables a reviewer to verify not only the results but the reasoning path that produced them.

The scenario catalog (S1–S9) was designed to stress-test the controller across mechanics, drive, and gas exchange perturbations. The framework successfully identified operating regions in which the AWP reaches its performance limits – specifically under severe gas exchange impairment (S6, S7) and stiff-lung mechanics (S2). Crucially, these failure modes were identified without exposing any patient to the associated risks.

This is precisely the use case envisioned by ASME V&V 40 for preclinical computational evaluation of closed-loop functions.

The occurrence of inspiratory pressure oscillations in 11.1% of simulated cases demonstrates that the framework is capable of reproducing emergent closed-loop dynamics that are not visible at the individual-breath level but arise from the interaction of the controller logic, patient physiology, and model dynamics over multiple adjustment cycles. The ability to detect such behavior in silico – and to compare its frequency with clinical observations – demonstrates a key capability of the PDM-based approach that simpler, open-loop simulation frameworks cannot provide.

The Spearman correlation analysis identified disease state, sedation level, and initial inspiratory pressure as dominant determinants of AWP performance, while demographic variables showed only secondary effects. This result is directly actionable at the framework level: it defines the parameter regions where a second-phase, higher-resolution ISCT should focus its sampling effort.

The comparison with retrospective clinical data serves as a framework-level credibility validation, not as a clinical evaluation of SmartCare®/PS. The qualitative alignment of performance trends – including the relative difficulty of lung-protective tidal volume control and the comparable oscillation rate (11.1% vs. 15.5%) – confirms that the PDM-based ISCT produces outputs that are consistent with real-world observations. The quantitative discrepancy in absolute TiT values is expected and explicable by the controlled simulation environment (absence of real-world confounders such as leakage, pain, or medication changes) and serves as a basis for identifying future model extensions rather than as a validation failure.

Collectively, these observations confirm that the framework is fit for its stated context of use: preclinical design evaluation and iterative refinement of automated ventilator functions prior to clinical investigation.

4.4 Limitations and Future Work

While the presented framework demonstrates a viable pathway for credibility-aware in silico clinical trials in mechanical ventilation, its current implementation reflects a methodological trade-off between complexity, computational feasibility, and clinical realism. The model, as presented, does not account for multi-morbidity, psychological factors such as pain or anxiety, or the dynamic interplay of systemic physiological changes over time. Factors that are known to significantly influence patient-ventilator interaction and ventilatory control in real-world settings. Furthermore, the simulation duration of 2000 seconds limits the assessment of long-term adaptation, physiological drift, or the cumulative effects of prolonged ventilatory support. These omissions are not inherent flaws of the framework itself, but rather deliberate simplifications made to enable a first demonstration of the methodological approach under controlled conditions. The current instantiation serves as a proof-of-concept for the framework methodology, not as a final credibility assessment of any specific device. The framework’s true potential lies in its extensibility: future iterations will move beyond isolated perturbations toward multi-event clinical scenarios that simulate disease progression, such as the evolution from acute respiratory distress to sepsis or recovery. Integrating real-world ICU data for cohort validation and model refinement will further enhance

the framework’s predictive power. Moreover, the inclusion of patient-specific risk profiling and adaptive sampling strategies will allow for targeted, high-resolution testing in safety-critical regions identified during initial screening. Ultimately, the framework is not a static tool, but a dynamic methodological platform that evolves from broad exploration to focused investigation. It is enabling a risk-informed approach to preclinical evaluation that aligns with both scientific rigor and regulatory expectations.

5 Conclusion

This work addresses a methodological gap in the preclinical evaluation of automated ventilator functions: the absence of an integrated, end-to-end ISCT pathway that coherently couples respiratory mechanics, gas exchange, respiratory drive, device logic, and measurement latencies while explicitly aligning with regulatory credibility frameworks. The proposed PDM-based ISCT framework closes this gap through three concrete contributions.

Contribution 1 – ASME V&V 40-aligned ISCT workflow

The framework enforces a top-down design logic in which all downstream decisions (model complexity, cohort composition, scenario catalog, and analysis methods) are constrained by the originally defined question of interest and context of use. This ensures full traceability from clinical question to quantitative conclusion, a precondition for regulatory use. Credibility activities are distributed across all workflow steps rather than confined to a single validation phase, reflecting the ASME V&V 40 principle that model credibility is a continuous property.

Contribution 2 – Virtual cohort and scenario methodology

The representative-edge-case stratification principle and hybrid LHS/uniform sampling strategy are generic framework elements that are independent of the specific device under test. They enable systematic population-level characterization and stress-testing of controller robustness at physiological boundaries that are under-represented in conventional trials.

Contribution 3 – Framework validation through case study

The case study of the automated weaning protocol confirmed that the framework generates decision-relevant evidence under real-world complexity. All five questions of interest were successfully addressed: the scenario catalog identified the conditions under which the AWP reaches its performance limits. Edge-case strata revealed the fundamental trade-off between CO₂ clearance and lung protection. Cohort heterogeneity enabled phenotype-specific characterization of patient-ventilator asynchrony and the framework reproduced emergent controller behavior (P_{insp} oscillations in 11.1% of simulations, compared to 15.5% in clinical data) without clinical risk exposure. Influence analysis identified disease state, sedation level, and initial inspiratory pressure as dominant performance determinants, providing a concrete basis for targeted cohort refinement in subsequent, higher-resolution trials.

The qualitative alignment with clinical comparator data confirms that the PDM-based ISCT produces outputs consistent with real-world observations, supporting the framework’s credibility for its stated context of use. Quantitative discrepancies are expected and informative, as they define the model extensions needed for future iterations.

The framework is device-agnostic and directly transferable to other automated ventilation strategies and closed-loop functions in related therapeutic domains. Future work will extend the framework toward multi-event disease progression scenarios, correlated multi-morbidity cohorts, and adaptive sampling in the critical parameter regions identified by the influence analysis, thereby progressing from broad exploratory screening toward high-resolution, statistically powered preclinical evaluation.

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

All authors have read and approved the final manuscript and consent to its publication.

Availability of supporting data

The in silico data used in this study were generated using a previously published model [<https://doi.org/10.1016/j.cmpb.2026.109329>]. The clinical data analyzed during the current study were collected with approval from the Ethics Committee of the Medical Faculty of the Christian-Albrechts-University of Kiel (approval no. D411/13). Due to ethical and data protection restrictions, the clinical data are not publicly available and cannot be shared.

Competing interests

Not applicable.

Funding

C. Hennigs, G. Männel, F. Bilda, H. Selpien, D. Schädler and P. Rostalski were supported by the German Federal Ministry of Research, Technology and Space (BMFTR) through the *CMS4Vent* Project (FKZ: 03LW0469K, 03LW0470, 03LW0471). C. Hennigs, G. Männel, F. Bilda and P. Rostalski were funded by the European Union - European Regional Development Fund (ERDF), the Federal Government and Land Schleswig Holstein, Project No. 12524009, and No. 12422005.

Table A1: Patient specific model parameter for chronic obstructive pulmonary disease (COPD) classified according to GOLD grading [27]. E_L : Lung elastance, R_1 : Inspiratory airway resistance, R_2 : Expiratory airway resistance, f_s : Shunt fraction, f_{VD} : Pulmonary dead space fraction, θ_{chem} : Set point resp. center for P_aCO_2 [mmHg]. calc: precalculated value based on patient demographics.

ID	E_L	R	f_s	f_{VD}	θ_{chem}	Literature
GOLD I	$E_{L,calc}/1.28$	$R_2 = R_{1,calc}$	0.06	0.06	40	[45]
GOLD II	$E_{L,calc}/1.73$	$R_2 = R_{1,calc} \cdot 3.17,$ $R_1 = R_{1,calc} \cdot 1.6$	0.12	0.12	40	[45]
GOLD III	$E_{L,calc}/3.8$	$R_2 = R_{1,calc} \cdot 5,$ $R_1 = R_{1,calc} = R_1 \cdot 2$	0.26	0.26	41	[45, 46]
GOLD IV	$E_{L,calc}/5.9$	$R_2 = R_{1,calc} \cdot 6.5,$ $R_1 = R_{1,calc} = R_1 \cdot 2.5$	0.37	0.37	43	[45, 46]

Authors' contributions

C.H., G.M., F.B., and P.R. conceived and designed the study. C.H. acquired the data, developed the software, performed the primary data analysis, and drafted the manuscript. H.S., D.S., and M.I. contributed to data analysis. M.I. and H.S. contributed to the interpretation of the results. M.I., H.S., and D.S. provided clinical expertise. D.S., P.R., G.M., and F.B. supervised the study. D.S., P.R., and F.B. contributed to project administration. P.R. provided resources. P.R., F.B., D.S., C.H., and G.M. acquired funding. All authors critically revised the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

Acknowledgements

To improve grammar, style, and overall readability of the manuscript, the authors used the following AI-based tools: DeepL Translator (DeepL SE, Cologne, Germany) and Fh-Genie (Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Munich, Germany). The authors have carefully reviewed all generated content and take full responsibility for the content of this work.

Appendix A

A.1 Virtual Cohort Characteristics

Tables A3–A2 summarize the patient-specific model parameter equations derived from demographic and anthropometric features (Table A3) as well as the disease-specific parameter adaptations for COPD (Table A1) and ARDS (Table A2), which together define the physiological basis for virtual cohort generation.

Table A2: Patient specific model parameter for ARDS classified according to the Berlin definition [28]. E_L : Lung elastance, f_s : Shunt fraction, f_{V_D} : Pulmonary dead space fraction.

ID	Parameter for non-linear E_L	f_s	f_{V_D}	Literature
Mild ARDS	$a_i = 2.11, b_i = 3.84, c_i = 19.05, d_i = 3.69,$ $a_e = 1.99, b_e = 3.91, c_e = 13.75, d_e = 4.17$	0.21	0.27	[47–49]
Moderate ARDS	$a_i = 1.76, b_i = 3.92, c_i = 23.11, d_i = 3.81,$ $a_e = 1.63, b_e = 3.95, c_e = 14.81, d_e = 4.20$	0.29	0.30	[47–49]
Severe ARDS	$a_i = 1.50, b_i = 3.63, c_i = 32.11, d_i = 3.82,$ $a_e = 1.34, b_e = 3.49, c_e = 17.75, d_e = 3.59$	0.40	0.38	[47–49]

A.2 Metric Definitions

A.2.1 Performance Metrics

For the preclinical evaluation of ventilatory performance, some QoI - where possible - were associated with target ranges: $V_T > 300$ mL, RR: 15 – 30 1/min, etCO_2 ≤ 55 mmHg based on SmartCare®/PS documentation [15]. In addition, lung-protective criteria were considered, namely a tidal volume $V_{T,LP}$ in the range of 6 – 8 ml/kg ideal body weight (IBW) and an plateau pressure $P_{\text{plat},LP} \leq 30$ cmH₂O [30].

To quantify performance, several key parameters are defined. Let $\text{QoI}(t)$ denote the value of the quantity of interest at time t , with T_{sim} representing the total simulation duration. The target range is defined by lower and upper bounds, Q_{low} and Q_{high} , respectively. These values are derived from clinical guidelines and device specifications (e.g., SmartCare®/PS documentation [15]). The following performance metrics are computed based on these definitions: time in target (TiT), overshoot, and longest excursion time (LET), which are used to assess the stability, accuracy, and robustness of the automated weaning protocol across diverse patient phenotypes and clinical scenarios. The first performance metric in this study is the time in target (TiT), which quantifies the fraction of the simulation duration during which a given quantity of interest remains within its predefined physiological target range $[Q_{\text{low}}, Q_{\text{high}}]$.

Time in target is defined as

$$\text{TiT} = \frac{1}{T_{\text{total}}} \int_0^{T_{\text{total}}} \mathbb{I}(Q_{\text{low}} \leq \text{QoI}(t) \leq Q_{\text{high}}) dt \cdot 100\%, \quad (\text{A1})$$

where T_{total} is the total simulation duration and $\mathbb{I}(\cdot)$ denotes the indicator function, equal to 1 if the condition is satisfied and 0 otherwise. Overshoot quantifies the maximum exceed of the upper target limit, expressed relative to the upper bound

$$\text{Overshoot} = \frac{\max(\text{QoI}(t)) - Q_{\text{high}}}{Q_{\text{high}}} \cdot 100\%. \quad (\text{A2})$$

Table A3: Patient-specific model parameters with standard deviation. m: Male, f: Female, IBW: Ideal body weight, BMI: Body mass index, REE: Resting energy expenditure, PAL: PAL-factor considers the individual activity level, TLCO: Transfer factor.

Description	Equation	Literature
Lung compliance [L/mbar]	$C_{L,m} = ((0.2891 + 0.0011 \cdot \text{age}) \cdot 0.76) \cdot \text{height}, \sigma = 0.0411 + 0.001 \cdot \text{age},$ $C_{L,f} = ((0.2891 + 0.0011 \cdot \text{age}) \cdot 0.76 \cdot 0.84) \cdot \text{height}, \sigma = 0.0411 + 0.001 \cdot \text{age}$	[32–34] [32–35]
Chest wall compliance [L/mbar]	$C_{CW,m} = (0.2439 - 0.0016 \cdot \text{age}) \cdot \text{height}, \sigma = 0.0876 - 0.0009 \cdot \text{age},$ $C_{CW,f} = (0.2133 - 0.0014 \cdot \text{age}) \cdot \text{height}, \sigma = 0.0408 - 0.0003 \cdot \text{age}$	[32, 35, 36] [35, 36]
Airway resistance [mbar/L/s]	$R = 2.55 \cdot e^{(0.03 \cdot \text{BMI})}$	[32, 37]
Cardiac output [L/min]	$Q_m = 7.2638 - 0.0265 \cdot \text{age}, \sigma = 1.8723 + 0.0111 \cdot \text{age},$ $Q_f = 5.6974 - 0.0190 \cdot \text{age}, \sigma = 1.8150 - 0.0140 \cdot \text{age}$	[38, 39] [38, 39]
Metabolic rate O ₂ [L/min]	$MP_{t,O_2} = ((\text{REE}_i/4.83)/(24 \cdot 60)) \cdot \text{PAL}$ with $i = m, f$ and $\text{PAL} = 0.95$	[38, 40, 41]
Metabolic rate CO ₂ [L/min]	$MP_{t,CO_2} = MP_{t,O_2} \cdot \text{RQ}$ with $\text{RQ} = 0.84$	[35, 38]
Diffusion capacity O ₂ [L/(s·mmHg)]	$D_{O_2} = (D_{CO} \cdot ((10^{-3} \cdot 22.4)/(60 \cdot 7.5))) \cdot \text{TLCO}, \sigma = (0.05 - D_{CO})/1.96$	[35, 38, 42]
Anatomical dead space [mL]	$V_D = 2.2 \frac{\text{mL}}{\text{kg}} \cdot \text{IBW}$	[43]
Functional residual capacity [mL]	$\text{FRC} = 30 \frac{\text{mL}}{\text{kg}} \cdot \text{IBW}$	[43]
Maximum inspiratory pressure [mmHg]	$\text{MIP}_m = 98.60 + 1.18 \cdot \text{age} + 0.76 \cdot \text{BMI} - 0.02 \cdot \text{age}^2,$ $\text{MIP}_f = 61.48 + 0.66 \cdot \text{age} + 1.55 \cdot \text{BMI} - 0.01 \cdot \text{age}^2$	[44] [44]
Maximum expiratory pressure [mmHg]	$\text{MEP}_m = 58.11 + 3.71 \cdot \text{age} + 2.64 \cdot \text{BMI} - 0.04 \cdot \text{age}^2,$ $\text{MEP}_f = 74.75 + 1.67 \cdot \text{age} + 1.75 \cdot \text{BMI} - 0.02 \cdot \text{age}^2$	[44] [44]

Longest excursion time (LET) is defined as the maximum continuous duration during which the QoI remains outside the target range

$$\text{LET} = \max\{\Delta t_{\text{excursion}} \mid \text{QoI}(t) \notin [Q_{\text{low}}, Q_{\text{high}}]\}. \quad (\text{A3})$$

A.2.2 Asynchrony Detection

The evaluation of the in silico simulations regarding patient-ventilator interaction and asynchrony is based on the work of Sauer et al. [29]. Synchronous triggers, delayed triggers, ineffective triggers, auto triggers, and double effort are evaluated, and the percentage of the total number of breaths is determined.

A.2.3 Inspiratory Pressure Oscillations

To assess the stability of the automated weaning protocol (AWP), inspiratory pressure (P_{insp}) time series were analyzed for oscillatory behavior induced by the controller's periodic adjustment of support levels. A k-means clustering algorithm was applied to segment each time series into two distinct states: a low- P_{insp} state (baseline support) and a high- P_{insp} state (adjusted support). State transitions and their durations were computed to identify oscillatory patterns. A transition was classified as a valid oscillation if it lasted between 120 and 300 seconds-matching the controller's predefined adjustment interval (every 2–5 minutes). The primary metric used to quantify oscillatory behavior was the number of valid oscillations per simulation run. Scenarios with more than three valid oscillations were classified as "oscillatory" to indicate sustained, controller-driven pressure cycling.

A.2.4 Composite Performance Score

Median values across all 400 virtual patients were first calculated for each quantity of interest (QoI) across the nine simulated scenarios. Each metric was normalized independently to a $[0, 1]$ scale using min-max normalization:

$$\tilde{M}_{i,j} = \frac{M_{i,j} - \min_{i'}(M_{i',j})}{\max_{i'}(M_{i',j}) - \min_{i'}(M_{i',j})}, \quad (\text{A4})$$

where $M_{i,j}$ is the median value of metric j in scenario i , and the minimum and maximum are taken over all scenarios $i' = 1, \dots, 9$. This ensures that the best-performing scenario for metric j receives a normalized score of 1, and the worst a score of 0. For each parameter $p \in \{V_T, \text{RR}, \text{etCO}_2, V_{T,\text{LP}}, P_{\text{plat},\text{LP}}\}$, the parameter-specific composite score $C_{p,i}$ was computed as the arithmetic mean of its normalized metrics:

$$C_{p,i} = \frac{1}{5} \sum_{j=1}^5 \tilde{M}_{i,j}^{(p)}, \quad (\text{A5})$$

where j indexes the five performance metrics (TiT, overshoot, LET, etc.) for parameter p . The overall composite score $C_{\text{overall},i}$ for scenario i was then calculated as the average of the five parameter-specific scores:

$$C_{\text{overall},i} = \frac{1}{5} \sum_p C_{p,i}. \quad (\text{A6})$$

This approach ensures that each metric contributes equally to the final score, while accounting for its intrinsic scale and range. The resulting score ranges from 0 (worst performance across all metrics) to 1 (best performance in all metrics), enabling a holistic, interpretable assessment of overall ventilatory performance.

A.2.5 Influence Analysis

Categorical variables (sex and disease categories) were encoded numerically. Spearman rank correlation coefficients ρ were then computed between all patient input variables and ventilation outcome measures, including V_T , RR, etCO_2 , $V_{T,LP}$, and $P_{\text{plat},LP}$ (each characterized by time in target, overshoot, and longest excursion), as well as P_{insp} oscillations and asynchrony indices.

A.3 Results of the Case Study in Detail

A.3.1 Performance Metrics, Composite Score, Asynchrony, and Inspiratory Pressure Oscillations

Performance metrics for tidal volume V_T , respiratory rate RR, end-tidal carbon dioxide etCO_2 and plateau pressure P_{plat} were computed for the nine exemplary simulated patient scenarios. Median values across 400 virtual patients were used for the analysis. Heatmaps displaying the time in target, overshoot, and longest excursion for each parameter are shown in Figure 3. Longest excursions were normalized to percentage of total simulation time ($T_{\text{total}} = 2000$ s). Mean values and standard deviation over all patients and scenarios are listed in Table 7.

For the acceptable working ranges, the time in target for V_T ranged from 49.4% (S2) to 84.5% (S7), indicating considerable variability in maintaining target tidal volumes across scenarios. Respiratory rate performance was markedly more stable, with TiT values exceeding 84% in all scenarios. Overshoot for RR was generally moderate (14.3 – 27.4%), and excursions remained below 2%. End-tidal CO_2 was well maintained with TiT over 80% in most scenarios, except S7 and S6 where TiT dropped to 54.7% and 33.0%, respectively. Overshoot was negligible across scenarios except in S7 (16.9%), while excursions ranged between 1.2% and 6.9% of total simulation time. Again, S6 and S7 are the exception with 54.8% and 13.8% excursion.

For lung protective ventilation, time in target for $V_{T,LP}$ ranged from 41.0% (S7) to 59.2% (S2), indicating considerable variability in maintaining target tidal volumes across scenarios. Overshoot values for $V_{T,LP}$ varied widely, with S7 reaching 42.2%, and longest excursions extended up to 10.6% of total simulation time, reflecting transient periods of significant deviation from target volumes. For the lung-protective parameter $P_{\text{plat},LP}$, time in target ranged from 43.96% (S6) to 100.00% (S1 and S2), indicating substantial variability in maintaining plateau pressures within the desired range across scenarios. Overshoot values for $P_{\text{plat},LP}$ varied between 0.00% (S1 and S2) and 17.32% (scenario 6). The longest excursions above the target region lasted up to 29.6% of total simulation time (S6), reflecting transient periods of marked deviation from the prescribed plateau pressure.

Normalized composite scores were calculated for each parameter by integrating time-in-target, overshoot, and excursion metrics. An overall composite score was obtained by averaging across V_T , RR, etCO_2 , and the lung protective parameter $V_{T,LP}$ and $P_{\text{plat},LP}$ (Figure 4a). Scenario 9 achieved the highest overall score (0.81), indicating the most balanced performance across all parameters in a weaning-like setting. S9 is closely followed by S5 (0.81) and S8 (0.78). In contrast, S6 exhibited the lowest overall score (0.44), reflecting pronounced imbalances between individual components. Across

scenarios, RR generally contributed the largest share to the overall composite score, with several scenarios reaching near-maximal RR performance (e.g. S6–S8). However, this high RR contribution often masked deficiencies in other parameters. In particular, etCO_2 performance was consistently reduced in S6 and 7, and V_T scores were markedly low in these cases. S7 showed the strongest divergence between RR (0.89) and V_T (0.05) as well as etCO_2 (0.38), underscoring the dominant impact of tidal volume excursions on the overall composite score. Conversely, S5 and S9 demonstrated more uniform contributions across all parameters, resulting in higher overall composite performance.

Asynchrony profiles revealed substantial scenario-dependent variability (Figure 4b). Delayed triggers were prevalent in S1, S4, and S5, whereas ineffective triggers dominated S2. Other asynchronous events, including double effort and auto-trigger occurred sporadically with scenario-specific patterns.

Regarding the P_{insp} oscillation (max. 4 mbar) analysis, out of all analyzed simulations, 11.05 % exhibited oscillatory behavior.

Overall, the results indicate that V_T , RR and etCO_2 were effectively maintained within target ranges. Scenario-dependent differences highlight the influence of ventilator settings and patient characteristics on ventilation performance and patient-ventilator synchrony.

A.3.2 Influence Analysis

Spearman correlation analysis was conducted (Figure A1). Age, height, BMI, and $F_i\text{O}_2$ exhibited generally weak correlations with all output parameters ($-\rho \leq 0.27$), indicating minimal monotonic influence. Age displayed at most a modest association with longer lung-protective tidal volume excursions ($\rho = 0.27$). Sex was moderately correlated with tidal volume time in target ($\rho = 0.38$) and negatively correlated with V_T overshoot ($\rho = -0.24$), suggesting a modest impact of patient sex on tidal volume control. Among ventilator settings, baseline inspiratory pressure and PEEP exhibited the most pronounced associations. Higher initial P_{insp} was correlated with reduced time in target for $P_{\text{plat,LP}}$ ($\rho = -0.16$), longer RR excursions ($\rho = 0.28$), and increased ineffective triggering ($\rho = 0.28$) and double efforts ($\rho = 0.26$). PEEP showed a similar pattern, with a moderate negative correlation with $P_{\text{plat,LP}}$ TiT ($\rho = -0.26$) and positive correlations with $P_{\text{plat,LP}}$ overshoot and longest excursion ($\rho = 0.20$ and 0.21 , respectively). The off-cycling sensitivity α and flow trigger β were modestly correlated with asynchrony metrics, particularly ineffective triggering ($\rho \approx 0.21$ for both) and double efforts ($\rho = 0.16$ and 0.15 for α and β , respectively), and showed weak-to-moderate relationships with lung-protective pressure metrics ($-\rho \leq 0.16$).

Deeper sedation showed the strongest correlations with reduced asynchronous events. Negative correlations were observed between ρ_{sed} and key asynchrony metrics, including double effort ($\rho = -0.41$). The sedation level also showed moderate associations with lung-protective tidal volume TiT ($\rho = 0.31$) and shorter RR excursions ($\rho = -0.25$). COPD was positively correlated with V_T TiT ($\rho = 0.25$) and $V_{T,LP}$ TiT ($\rho = 0.15$) and negatively with overshoot parameters for RR and etCO_2 ($\rho = -0.22$ and -0.15 , respectively), suggesting more stable controller performance in obstructive

phenotypes. In contrast, ARDS was associated with impaired lung-protective control, characterized by reduced $P_{\text{plat,LP}}$ TiT ($\rho = -0.34$) and $V_{\text{T,LP}}$ TiT ($\rho = -0.19$), together with higher $P_{\text{plat,LP}}$ overshoot ($\rho = 0.36$) and prolonged $V_{\text{T,LP}}$ and $P_{\text{plat,LP}}$ excursions ($\rho = 0.35$ and 0.32 , respectively). Other diagnostic groups (postoperative, neuromuscular, lung edema) showed only weak associations with most outcome metrics ($-\rho \leq 0.19$). The occurrence of P_{insp} oscillations was primarily related to disease category, showing a positive correlation with ARDS ($\rho = 0.25$) and a negative with COPD ($\rho = -0.26$).

These findings could provide a basis for subsequent, more targeted analyses: identified risk groups such as severe ARDS patients with high initial inspiratory pressure or deeply sedated patients. This could be investigated in greater detail by refining the virtual cohort through denser sampling in the corresponding parameter regions, thereby enabling a higher-resolution characterization of controller performance and failure modes within these critical subpopulations.

A.3.3 Detailed Interpretation per Question of Interest

The following discussion first addresses the four predefined questions of interest, examining how the results relate to (i) acceptable working ranges, (ii) lung-protective ventilation, (iii) patient-ventilator asynchrony, and (iv) the inspiratory pressure oscillations. Ventilatory performance was quantified across nine simulated clinical scenarios using time in target, overshoot, longest excursions, asynchrony profiles, P_{insp} oscillations, and normalized composite scores.

Question of Interest 1: Acceptable Working Ranges

For the acceptable working ranges of the automated weaning protocol, tidal volume performance was generally in the mid-to-upper range of the predefined TiT interval, with only limited boundary violations. The relatively short longest excursions (typically corresponding to only a few breaths or short phases of boundary crossing; e.g. $\approx 4\%$ of $2000 \text{ s} \approx 80 \text{ s}$) suggest that deviations were transient rather than sustained. S2 exhibited the lowest TiT for V_{T} among the acceptable working-range scenarios, which is consistent with the altered respiratory mechanics in this case (increased lung elastance) that make it more difficult for the controller to maintain the target volume. Especially when ineffective triggering is present, as in this case. A similar pattern is observed in S4, where respiratory failure limits the patient’s ability to compensate and P_{insp} adaptation requires several adjustment cycles before a new steady state is reached.

The simulated respiratory rate demonstrated high stability across all scenarios, with high TiT, moderate overshoot, and very short excursion times. Similarly, etCO_2 was well maintained, with near-perfect TiT in most scenarios and minimal overshoot. The target ranges for these parameters were chosen relatively wide (RR: $15 - 30 \text{ 1/min}$, $\text{etCO}_2 \leq 55 \text{ mmHg}$), which likely contributed to the observed stability. For etCO_2 , performance was generally good, except in S6 (increased shunt fraction f_s) and S7 (increased dead space V_{D}). Both pathophysiological alterations elevate etCO_2 and therefore challenge the AWP. The higher ventilatory demands lead to increased P_{insp} support. The minute ventilation emerges from the interaction between RR and V_{T} ,

which are modulated differently by the AWP: P_{insp} directly influences V_T on a breath-by-breath basis, and together with RR this determines MV, which in turn drives changes in etCO_2 . As MV increases, CO_2 levels fall, reducing chemoreflex drive and thereby indirectly slowing RR. Conversely, low MV leads to CO_2 retention, increased drive, and a reflexive rise in RR. These results indicate that the current controller is fundamentally well suited to maintaining V_T , RR, and etCO_2 within their predefined working ranges.

Question of Interest 2: Lung-Protective Ventilation

The results suggest that the AWP tends to operate V_T near the upper lung-protective limit, so that even modest increases in inspiratory effort can push individual breaths beyond the desired range. The observed oscillations in P_{insp} are consistent with this behavior and indicate that the controller intermittently switches between adjusting and holding P_{insp} at its bounds in an attempt to balance adequate ventilation and lung protection. This implies that strictly lung-protective V_T control is difficult to fully automate without additional feedback (e.g. direct monitoring of patient effort or alveolar strain) and that clinician oversight remains essential, particularly when V_T persistently deviates from target. In this context, hybrid automation - where the algorithm adjusts P_{insp} while clinicians supervise V_T and intervene as needed - likely reflects clinical practice better than fully autonomous control. A more aggressive or faster V_T controller could, in principle, keep V_T within target ranges more effectively, but only at the cost of increased control complexity.

In this *in silico* clinical trial, the challenge of lung-protective V_T control arises from several interacting factors. First, the target range for lung-protective ventilation (6 – 8 mL/kg) is narrower than the acceptable working range, leaving less tolerance for transient deviations. Second, V_T depends on compliance, resistance, and respiratory drive - parameters with substantial inter-patient variability and scenario-specific changes. Third, P_{insp} is only updated every 120 seconds, which may be too slow to track rapid V_T excursions driven by fluctuations in patient effort. Finally, overshoots of $V_{T,\text{LP}}$ of up to $\sim 40\%$ suggest either classical control delay effects (overshoot after an increase in P_{insp} followed by undershoot during correction) or highly variable patient effort. The relatively short excursion times support the latter interpretation, indicating predominantly transient boundary crossings rather than a sustained loss of controller. For plateau pressure, the lung-protective target of $P_{\text{plat}} \leq 30$ mbar was met for the majority of scenarios for most of the simulation time, with TiT for $P_{\text{plat,LP}}$ approaching 100% in S1 and S2 and remaining high in most of the remaining scenarios. This indicates that, under standard conditions, the AWP successfully maintains plateau pressure well below injurious levels. However, S3, S6, and S7 revealed important limitations. In S3, increased airway resistance prolongs the respiratory time constant, making it more difficult to deliver the target volume within the available inspiratory time. As a result, P_{insp} must be raised substantially, which brings P_{plat} closer to or beyond the 30 mbar threshold and increases overshoot and excursion duration. In S6 and S7, both of which feature pronounced impairments of gas exchange (increased shunt and increased dead space, respectively), the resulting rise in chemoreflex drive leads to high respiratory rates (as also evident in Figure 3) and a strong tendency

toward higher inspiratory pressures. With a too high RR, the AWP increases P_{insp} , but in the most severely impaired cases this is not sufficient to fully normalize gas exchange without transiently exceeding the plateau pressure limit. Consequently, $P_{\text{plat,LP}}$ TiT is reduced and overshoot and longest excursions increase in these scenarios.

Together, these observations highlight a fundamental tension between gas exchange optimization and strict lung protection: in scenarios with severely impaired mechanics or gas exchange, maintaining etCO_2 and RR within target ranges may require P_{insp} levels that conflict with conservative $V_{\text{T,LP}}$ and $P_{\text{plat,LP}}$ goals.

Questions of Interest 1 and 2: Composite Score

Normalized composite scores of the simulation results, integrating TiT, overshoot, and excursion metrics, provided a concise summary of overall ventilatory performance. An overall composite score was obtained by averaging across V_{T} , RR, etCO_2 , and the lung-protective parameters $V_{\text{T,LP}}$ and $P_{\text{plat,LP}}$ (Figure 4a). S9 achieved the highest overall score (0.81), indicating the most balanced performance across all components in a weaning-like setting. S5 (0.81) and 8 (0.78) showed similarly high scores, reflecting comparatively robust control of both gas exchange and lung-protective targets under these conditions. In contrast, S6 exhibited the lowest overall score (0.44), highlighting pronounced imbalances between individual parameters under severe shunt conditions. Across scenarios, RR generally contributed the largest share to the overall composite score, with several scenarios reaching near-maximal RR performance (e.g. S6–S8). However, this strong RR contribution frequently masked deficiencies in other parameters. In particular, etCO_2 performance was consistently reduced in S6 and S7, and V_{T} -based scores (both V_{T} and $V_{\text{T,LP}}$) were markedly low, indicating that adequate control of rate alone is insufficient to ensure lung-protective ventilation when gas exchange is severely impaired. S7 showed the most pronounced divergence between RR (0.89) and V_{T} (0.05) as well as etCO_2 (0.38), underscoring the dominant impact of tidal volume excursions and CO_2 dysregulation on overall performance. Conversely, S5 and S9 demonstrated more uniform contributions from V_{T} , RR, etCO_2 , $V_{\text{T,LP}}$, and $P_{\text{plat,LP}}$, resulting in higher composite scores and indicating that the AWP is particularly well suited to gradual or abrupt changes in sedation level, as typically encountered during weaning. Overall, these findings emphasize that the quality of automated ventilatory support is strongly scenario-dependent and that a high RR performance alone does not guarantee balanced control of gas exchange and lung-protective targets.

Question of Interest 3: Asynchrony

Analysis of asynchronous events of the simulation results revealed frequent delayed and ineffective triggers across scenarios. Scenario-specific patterns of double effort and auto-trigger suggested that patient characteristics and ventilator settings interact to influence asynchrony. Clinically, such asynchronies can compromise ventilation efficiency, increase work of breathing, and potentially contribute to patient discomfort or ventilator-induced lung injury [31].

When comparing to clinical literature [29, 31], delayed triggers were increased at rates up to $\approx 35\%$ across all scenarios. This is understandable given the wide choice of flow triggers β in this simulated cohort. The higher the trigger threshold is set, the more

likely it is that the trigger will be delayed or ineffective. Ineffective triggers simulated at rates of 5 – 10 % are again increased in comparison literature data [29, 31, 50]. In scenario 2, most triggers were ineffective, underlining the difficulty of ventilation of a very stiff lung, as e.g. typical for severe ARDS [28].

Question of Interest 4: Inspiratory Pressure Oscillations

The results demonstrate that oscillatory behavior does emerge in silico allowing for a systematic analysis. The clinical significance of these P_{insp} oscillations remains unclear. Recent clinical studies on noisy pressure support ventilation - which introduces breath-to-breath variability - suggest that moderate variability is safe and may even improve patient-ventilator synchrony and respiratory mechanics [51, 52]. Importantly, noisy PSV varies the pressure on a per-breath basis, whereas the oscillations observed in the present study result from slower, controller-driven switching between discrete P_{insp} levels every 2–5 minutes. Consequently, the physiological implications of slower oscillations may differ from those of breath-to-breath variability, and negative effects cannot be ruled out.

Question of Interest 5: Dominant Parameters

Overall, clinically plausible trends were observed, with disease category, sedation level, and initial inspiratory pressure exerting the strongest effects, while demographic variables showed weaker but interpretable influences.

Sex-related differences were primarily reflected in tidal volume performance. Lower V_T time-in-target in female patients is consistent with higher lung elastance associated with smaller lung volumes [35]. A similar but weaker effect was observed for height, indicating that anthropometric scaling rather than sex itself drives these differences. Body mass index showed a mild influence on respiratory rate and asynchrony metrics, in line with increased elastance and airway resistance at higher BMI [37], which can impair patient-ventilator interaction.

Initial inspiratory pressure P_{insp} was a key determinant of performance, particularly for V_T and RR time-in-target. High initial P_{insp} values increased the likelihood of starting outside target ranges, leading to larger overshoots and longer correction times. This was also associated with increased asynchrony and P_{insp} oscillations, reflecting interactions between ventilatory support level, patient drive, and controller dynamics. Trigger parameters (α , β) showed expected but overall modest effects on ineffective and double triggering, likely because they were not changed during the simulations. In contrast, $F_i\text{O}_2$ had no measurable impact on performance metrics, as hypoxic drive was not reached in the simulated scenarios.

Sedation level substantially affected ventilation outcomes by modulating respiratory drive. Deeper sedation influenced V_T , RR, and etCO_2 , and was associated with altered asynchrony patterns, such as missed or delayed triggering. Underlying disease states showed the strong effects as well. ARDS and COPD most negatively impacted time-in-target metrics and etCO_2 control, consistent with their pronounced effects on lung compliance, airway resistance and gas exchange.

Inspiratory pressure oscillations represent an important emergent behavior of the system. Their occurrence in the in silico cohort was comparable to clinical observations, indicating that relevant controller instabilities were reproduced.

In summary, ventilatory performance was primarily determined by disease state, sedation level, and initial inspiratory pressure, while demographic factors played a secondary role mediated through lung mechanics. The alignment of simulated trends with clinical expectations supports the validity of the in silico cohort. At the same time, reduced variability compared with clinical data highlights the controlled nature of the model and points to future extensions, such as increased physiological variability and measurement noise, to further enhance realism.

The present analysis represents an initial, broad screening of the parameter space across a heterogeneous virtual cohort. Building on the identified risk factors and dominant input–output associations, subsequent analyses should employ larger, more densely sampled cohorts in the critical parameter regions – such as severe ARDS with high initial inspiratory pressure or low sedation levels – to enable a higher-resolution characterization of controller performance and failure modes within these vulnerable subpopulations.

Last, it should be noted that the Spearman correlation analysis was performed across all nine scenarios combined. As a result, the reported coefficients reflect global, scenario-aggregated trends and may conflate regime-specific behavior (e.g. shunt-dominated versus dead-space-dominated conditions or changes in respiratory drive versus mechanics). While this global analysis is useful for identifying dominant monotonic associations and ranking influential variables at the cohort level, it may obscure scenario-specific effects. A stratified correlation analysis per scenario could provide additional insight into how patient and ventilator characteristics influence performance under distinct pathophysiological conditions and will be the subject of future work.

A.4 Clinical Comparator Data: Detailed Analysis

The clinical dataset comprised recordings from 65 patients, totaling 347 individual measurement files. The number of recordings per patient ranged from 1 to 23, with a mean of 5.3 recordings per patient. The mean duration of a single recording was 995.3 ± 145.0 minutes, corresponding to a total analyzed data volume of 5225 hours. Regarding data availability, valid measurements were present in 77.8% of recordings for V_T , 90.8% for RR, 81.8% for $etCO_2$, and 90.5% for P_{insp} . Asynchrony detection was not possible, due to missing reference measurements of the patients muscle activity.

It should be explicitly noted that the AWP implemented in the ISCT framework represents a simplified abstraction of the SmartCare[®]/PS logic [15], derived from published protocol descriptions rather than from proprietary device firmware. Consequently, the in silico AWP and the clinically deployed system may differ substantially. Demographic and anthropometric characteristics of the clinical cohort and the in silico clinical trial population are summarized in Table 7. Mean age was slightly higher in the ISCT population (75.4 ± 15.7 years) than in the clinical cohort (70.3 ± 11.6 years). Sex distribution was similar (68% vs. 64% male), and average height and BMI were nearly identical (1.73 ± 0.10 m, 26.0 ± 5.2 kg/m² vs. 1.73 ± 0.09 m, 25.9

$\pm 6.6 \text{ kg/m}^2$). These results indicate that, despite a slightly older age distribution, the simulated population is demographically well aligned with the clinical cohort. No conclusions can be drawn regarding the distribution of underlying diagnoses, as clinical diagnostic data are not available for comparison.

Ventilatory performance metrics were analyzed in terms of time in target, overshoot, and longest excursion. For V_T , the clinical cohort exhibited a substantially higher TiT ($94.32 \pm 4.61 \%$) than the ISCT population ($69.74 \pm 11.30 \%$). Longest excursions were markedly longer in the clinical data ($1424 \pm 4748 \text{ s}$ vs. $108.0 \pm 35.60 \text{ s}$), indicating more prolonged deviations from the tidal volume target at the bedside. It should be noted, however, that the comparison of excursion duration between clinical and ISCT data is inherently limited, as clinical recordings often span several hours, whereas each simulation run covers only 2000 s.

For RR, both cohorts maintained high TiT (clinical: $97.9 \pm 21.2 \%$; ISCT: $95.7 \pm 4.1 \%$), but the ISCT data showed higher overshoot ($17.9 \pm 4.5 \%$ vs. $6.7 \pm 26.4 \%$) and substantially shorter excursions ($17.3 \pm 4.8 \text{ s}$ vs. $4046 \pm 17929 \text{ s}$).

For etCO_2 , simulations achieved a higher TiT ($93.2 \pm 19.7 \%$) compared to clinical recordings ($81.8 \pm 40.4 \%$), with both datasets showing negligible overshoot (clinical: $0 \pm 10.6 \%$; ISCT: $0 \pm 3.9 \%$).

Lung-protective performance was assessed using $V_{T,LP}$ and $P_{plat,LP}$. For $V_{T,LP}$, the ISCT cohort achieved a higher TiT ($51.3 \pm 6.2 \%$) than the clinical data ($38.54 \pm 23.28 \%$), with notably lower overshoot ($30.0 \pm 9.2 \%$ vs. $93.2 \pm 73.3 \%$). In contrast, $P_{plat,LP}$ TiT was high in both cohorts and slightly higher in the clinical data ($91.22 \pm 25.36 \%$ vs. $83.32 \pm 21.80 \%$ in ISCT). Overshoot remained relatively low for both ($9.47 \pm 23.11 \%$ clinical vs. $6.12 \pm 6.67 \%$ ISCT).

Inspiratory pressure oscillations occurred in 11.1 % of simulated recordings, compared to 15.5 % in the clinical data, indicating that the simplified implementation of airway pressure control in the ISCT framework captures oscillatory behavior of comparable magnitude to that observed in real patients.

The in silico cohort reproduces key ventilatory patterns observed in clinical data. Target adherence for tidal volume and respiratory rate is generally higher in simulations, however the trends in these metrics correspond well with the clinical measurements. Demographic characteristics of the in silico cohort, including age, sex, height, and body mass index, closely match those observed in the clinical dataset supporting the design and sampling of the digital patient cohort. This similarity supports the notion that the in silico population is representative of the clinical cohort in terms of basic anthropometric features, allowing for a meaningful comparison of ventilatory dynamics. It should be noted, however, that no conclusions can be drawn regarding disease distributions, as diagnostic information was not available for the clinical data. Some differences are explained by the controlled simulation environment. Real-life complications such as leakage, stress, pain, medication effects, or variations in sleep state are not modeled, leading to lower variability and shorter excursion duration in the in silico clinical trial. Simulation times are also shorter (2000 s versus hours or days in clinical recordings), limiting the comparability of absolute excursion duration. Notably, inspiratory pressure oscillations, which occurred in approximately 11.1 % of

simulated recordings compared to 15.5 % in clinical data, demonstrate that the ISCT captures even subtle, emergent behaviors of the ventilatory controller. This alignment suggests that the simplified implementation of the automatic weaning protocol is sufficient to elicit clinically relevant oscillatory behavior.

Despite these simplifications, the occurrence and characteristics of P_{insp} oscillations indicate that the model is capable of generating physiologically plausible dynamics. Incorporating additional sources of variability and realistic perturbations could further enhance the fidelity of the simulations. Overall, the ISCT successfully reproduces both general trends and emergent phenomena such as oscillatory behavior, providing a solid basis for *in silico* studies and future investigations of patient-specific ventilatory responses.

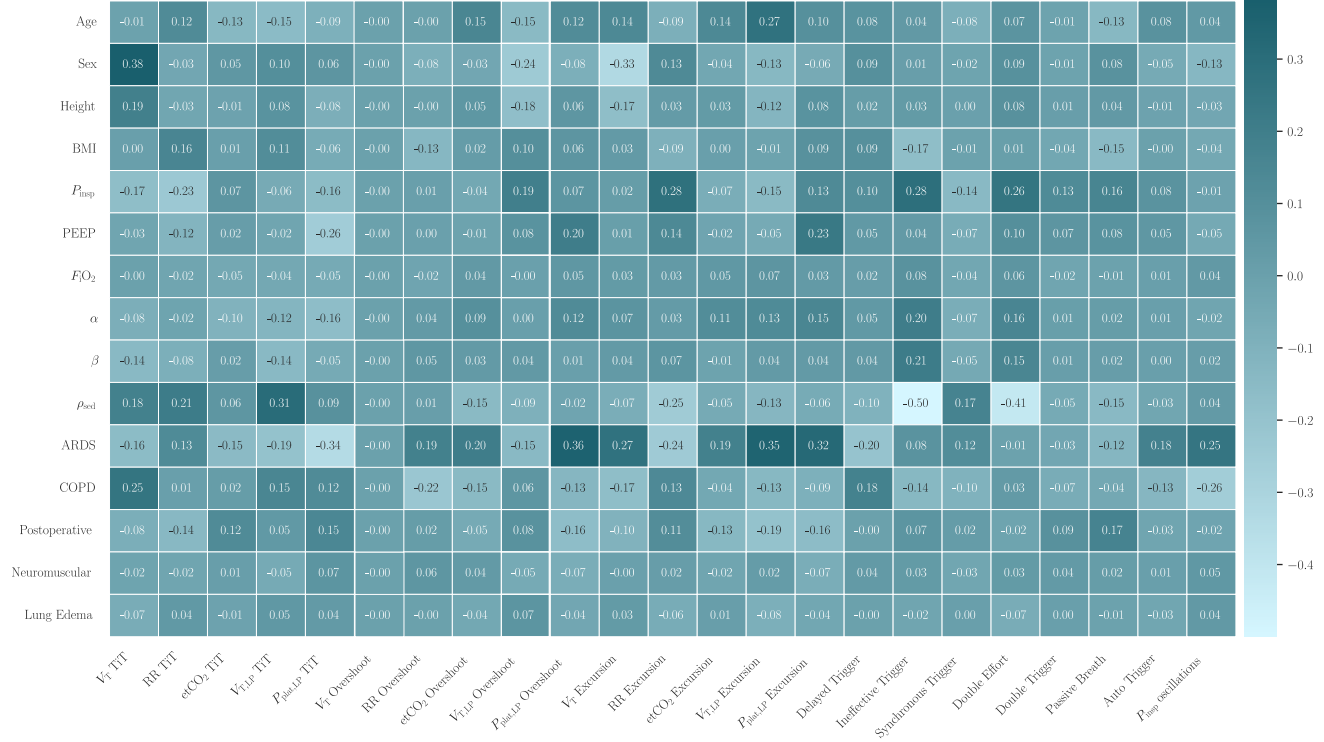


Fig. A1: Spearman correlation between patient characteristics and ventilation performance parameters. The heatmap shows Spearman correlation coefficients ρ between patient inputs - including demographics (age, sex, height and BMI), clinical parameters (inspiratory pressure P_{insp} , PEEP, inspiratory pressure of O_2 $F_{\text{I}}\text{O}_2$, off-cycling sensitivity α , flow-trigger β , and sedation level scaling factor ρ_{sed}), and disease categories (ARDS, COPD, postoperative, neuromuscular, lung edema) - and ventilation output parameters across 400 simulated patients under nine ventilatory scenarios. Positive correlations (red) indicate that higher input values are associated with higher output measures, while negative correlations (blue) indicate an inverse relationship.

References

- [1] F. Pappalardo et al., In silico clinical trials: concepts and early adoptions, *Briefings in Bioinformatics*, vol. 20, no. 5, pp. 1699–1708, 2019, doi: 10.1093/bib/bby043.
- [2] G. Karanasiou et al., Advancing in Silico Clinical Trials for Regulatory Adoption and Innovation, *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 4, pp. 2654–2668, 2025, doi: 10.1109/JBHI.2024.3486538.
- [3] The American Society of Mechanical Engineers. ASME V&V 40-2018 - Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices. 2018.
- [4] U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Devices and Radiological Health. Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions - Guidance for Industry and FDA Staff. 2023.
- [5] Virtual Physiological Human Institute. Virtual Physiological Human Institute. 2024. url: <https://www.vph-institute.org/> (visited on 03/18/2026).
- [6] Avicenna Alliance. Avicenna Alliance. 2024. url: <https://avicenna-alliance.com/> (visited on 03/18/2026).
- [7] M. Viceconti and L. Emili, (eds.) *Toward Good Simulation Practice: Best Practices for the Use of Computational Modelling and Simulation in the Regulatory Process of Biomedical Products*, 1st ed. 2024 ed. Cham: Springer Nature Switzerland, 2024, isbn: 978-3-031-48286-1 978-3-031-48284-7.
- [8] C. Y. S. Ang et al., Virtual patient with temporal evolution for mechanical ventilation trial studies: A stochastic model approach, *Computer Methods and Programs in Biomedicine*, vol. 240, pp. 107728, 2023, doi: 10.1016/j.cmpb.2023.107728.
- [9] D. Ratano et al., Lung- and diaphragm-protective strategies in acute respiratory failure: an in silico trial, *Intensive Care Medicine Experimental*, vol. 12, no. 1, pp. 20, 2024, doi: 10.1186/s40635-024-00606-x.
- [10] K. M. Förster et al., In silico numerical simulation of ventilator settings during high-frequency ventilation in preterm infants, *Pediatric Pulmonology*, vol. 56, no. 12, pp. 3839–3846, 2021, doi: 10.1002/ppul.25626.
- [11] L. Fresiello et al., Initial clinical validation of a hybrid in silico—in vitro cardiorespiratory simulator for comprehensive testing of mechanical circulatory support systems, *Frontiers in Physiology*, vol. 13, pp. 967449, 2022, doi: 10.3389/fphys.2022.967449.
- [12] Jacob Herrmann et al., Virtual patient model for evaluating automated inspired oxygen control, *Comput. Biol. Medicine*, 2025, doi:

10.1016/j.compbimed.2025.110505.

- [13] G. Männel et al., Optimizing mechanical power for lung protective ventilation – a simulation study, at - Automatisierungstechnik, vol. 73, no. 12, pp. 934–950, 2025, doi: 10.1515/auto-2025-0077.
- [14] C. Schranz et al., Model-Based Ventilator Settings in Pressure Controlled Ventilation, Biomedical Engineering / Biomedizinische Technik, 2013, doi: 10.1515/bmt-2013-4425.
- [15] A. Neumann and H. Schmidt, SmartCare/PS - Das automatisierte Weaning Protokoll, 2015.
- [16] C. Hennigs et al., Model Credibility Assessment of a Computational Respiratory System Model for Mechanical Ventilation, preprint on arXiv, doi.org/10.48550/arXiv.2607.06210.
- [17] C. Hennigs et al., A switching lung mechanics model for detection of expiratory flow limitation, at - Automatisierungstechnik, vol. 72, no. 5, pp. 417–428, 2024, doi: 10.1515/auto-2023-0206.
- [18] C. Hennigs, T. Becher, and P. Rostalski, Mathematical lung model for local gas exchange based on EIT-measurements, Current Directions in Biomedical Engineering, vol. 8, no. 2, 2022.
- [19] C. Hennigs et al., Patient-Ventilator Interaction - Development of a Mathematical Model of the Respiratory Center, Computer Methods and Programs in Biomedicine, pp. 109329, 2026, doi: 10.1016/j.cmpb.2026.109329.
- [20] C. Danielson et al., Credibility Assessment of Mathematical Physiological Respiratory Models used in In Silico Clinical Trials, In preparation, 2026.
- [21] D. Schädler. Non-invasive estimation of the work rate of breathing during automated weaning with SmartCare/PS: a prospective observational trial, Clinical data. 2010.
- [22] A. Garland et al., Epidemiology of critically ill patients in intensive care units: a population-based observational study, Critical Care, vol. 17, no. 5, pp. R212, 2013, doi: 10.1186/cc13026.
- [23] D. Garcia-Castellote et al., Evaluation of indirect measures of neural inspiratory time from invasive and noninvasive recordings of respiratory activity, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2017, pp. 341–344, 2017, doi: 10.1109/embc.2017.8036832.
- [24] A. S. Neto et al., Epidemiological characteristics, practice of ventilation, and clinical outcome in patients at risk of acute respiratory distress syndrome in intensive

- care units from 16 countries (PROVENT): an international, multi-centre, prospective study, *The Lancet Respiratory Medicine*, vol. 4, no. 11, pp. 882–893, 2016, doi: 10.1016/S2213-2600(16)30305-8.
- [25] A. Anzueto et al., Influence of body mass index on outcome of the mechanically ventilated patients, *Thorax*, vol. 66, no. 1, pp. 66–73, 2011, doi: 10.1136/thx.2010.145086.
- [26] Our World in Data. Human height across the world. 2025. url: <https://ourworldindata.org/human-height> (visited on 12/20/2025).
- [27] S. Mirza et al., COPD Guidelines: A Review of the 2018 GOLD Report, *Mayo Clinic Proceedings*, vol. 93, no. 10, pp. 1488–1502, 2018, doi: 10.1016/j.mayocp.2018.05.026.
- [28] Acute Respiratory Distress Syndrome: The Berlin Definition, *JAMA*, vol. 307, no. 23, 2012, doi: 10.1001/jama.2012.5669.
- [29] J. Sauer et al., Automated characterization of patient-ventilator interaction using surface electromyography, *Annals of Intensive Care*, vol. 14, no. 1, pp. 32, 2024, doi: 10.1186/s13613-024-01259-5.
- [30] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. S3-Leitlinie Invasive Beatmung und Einsatz extrakorporaler Verfahren bei akuter respiratorischer Insuffizienz. Version 2.0. S3-Leitlinie der Deutschen Gesellschaft für Anästhesiologie und Intensivmedizin (DGAI) und weiterer Fachgesellschaften, AWMF-Reg.Nr. 001-021. 2025.
- [31] A. W. Thille et al., Patient-ventilator asynchrony during assisted mechanical ventilation, *Intensive Care Medicine*, vol. 32, no. 10, pp. 1515–1522, 2006, doi: 10.1007/s00134-006-0301-8.
- [32] S. Merrath et al., A lung model considering age, sex and height: an in-silico study-lung mechanics, *Proceedings on Automation in Medical Engineering Proc AUTOMED*, vol. 2, No 1, 2023, doi: 10.18416/AUTOMED.2023.
- [33] C. Mittman et al., Relationship between chest wall and pulmonary compliance and age, *Journal of Applied Physiology*, vol. 20, no. 6, pp. 1211–1216, 1965, doi: 10.1152/jappl.1965.20.6.1211.
- [34] M. Mezidi and C. Guérin, Effects of patient positioning on respiratory mechanics in mechanically ventilated ICU patients, *Annals of Translational Medicine*, vol. 6, no. 19, pp. 384–384, 2018, doi: 10.21037/19858.
- [35] P. Haber, *Lungenfunktion und Spiroergometrie: Interpretation und Befunderstellung unter Einschluss der arteriellen Blutgasanalyse*, Vienna: Springer Vienna, 2013, isbn: 978-3-7091-1276-2 978-3-7091-1277-9.

- [36] M. Estenne, J. C. Yernault, and A. De Troyer, Rib cage and diaphragm-abdomen compliance in humans: effects of age and posture, *Journal of Applied Physiology*, vol. 59, no. 6, pp. 1842–1848, 1985, doi: 10.1152/jappl.1985.59.6.1842.
- [37] P. Pelosi et al., The effects of body mass on lung volumes, respiratory mechanics, and gas exchange during general anesthesia, *Anesthesia and Analgesia*, vol. 87, no. 3, pp. 654–660, 1998, doi: 10.1213/00000539-199809000-00031.
- [38] S. Merrath et al., A lung model considering age, sex and height: an in-silico study-gas exchange, *Proceedings on Automation in Medical Engineering Proc AUTOMED*, vol. 2, No 1, 2023, doi: 10.18416/AUTOMED.2023.
- [39] A. A. Luisada et al., Changes of cardiac output caused by aging: an impedance cardiographic study, *Angiology*, vol. 31, no. 2, pp. 75–81, 1980, doi: 10.1177/000331978003100201.
- [40] M. D. Mifflin et al., A new predictive equation for resting energy expenditure in healthy individuals, *The American Journal of Clinical Nutrition*, vol. 51, no. 2, pp. 241–247, 1990, doi: 10.1093/ajcn/51.2.241.
- [41] D. Mathias, *Fit und gesund von 1 bis Hundert*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, isbn: 978-3-662-56306-9 978-3-662-56307-6.
- [42] S. Stanojevic et al., Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians, *European Respiratory Journal*, vol. 50, no. 3, pp. 1700010, 2017, doi: 10.1183/13993003.00010-2017.
- [43] A. B. Lumb, *Nunn’s applied respiratory physiology*, Eighth ed. Edinburgh; New York: Elsevier, 2017, isbn: 978-0-7020-6294-0.
- [44] A. Lista-Paz et al., Maximal Respiratory Pressure Reference Equations in Healthy Adults and Cut-off Points for Defining Respiratory Muscle Weakness, *Archivos de Bronconeumología*, vol. 59, no. 12, pp. 813–820, 2023, doi: 10.1016/j.arbres.2023.08.016.
- [45] L. Jarenbäck et al., Flow-Volume Parameters in COPD Related to Extended Measurements of Lung Volume, Diffusion, and Resistance, *Pulmonary Medicine*, vol. 2013, pp. 1–10, 2013, doi: 10.1155/2013/782052.
- [46] R. J. Huijsmans et al., The clinical utility of the GOLD classification of COPD disease severity in pulmonary rehabilitation, *Respiratory Medicine*, vol. 102, no. 1, pp. 162–171, 2008, doi: 10.1016/j.rmed.2007.07.008.
- [47] M. Denai, M. Mahfouf, and G. H. Mills, Modeling and simulation of Electrical Impedance Tomography (EIT) on ventilated patients with ARDS lungs, pp. 1–6, 2008, doi: 10.1109/BIBE.2008.4696762.

- [48] L. Kadkhodai et al., Estimating the best fraction of inspired oxygen for calculation of PaO₂/FiO₂ ratio in acute respiratory distress syndrome due to COVID-19 pneumonia, *Journal of Research in Medical Sciences*, vol. 27, no. 1, pp. 38, 2022, doi: 10.4103/jrms.jrms-558-21.
- [49] R. H. Kallet et al., Lung Injury Etiology and Other Factors Influencing the Relationship Between Dead-Space Fraction and Mortality in ARDS, *Respiratory Care*, vol. 62, no. 10, pp. 1241–1248, 2017, doi: 10.4187/respcare.05589.
- [50] Q. Mulqueeny et al., Automatic detection of ineffective triggering and double triggering during mechanical ventilation, *Intensive Care Medicine*, vol. 33, no. 11, pp. 2014–2018, 2007, doi: 10.1007/s00134-007-0767-z.
- [51] P. M. Spieth et al., Short-term effects of noisy pressure support ventilation in patients with acute hypoxemic respiratory failure, *Critical Care*, vol. 17, no. 5, pp. R261, 2013, doi: 10.1186/cc13091.
- [52] Z. Jiang et al., Comparisons of respiratory mechanics in patients under different variations of noisy pressure support ventilation during ventilator weaning, *Journal of Thoracic Disease*, vol. 17, no. 10, pp. 8458–8466, 2025, doi: 10.21037/jtd-2025-78.