Project No. 101045956
Biomarker and AI-supported FX06 therapy to prevent progression from mild and moderate to severe stages of COVID-19

Deliverable 6.8

All relevant data and data-analysis results made available to a relevant data portal

WP 6 – Decision support models

Lead Participant	Tampereen Korkeakoulusäätiö SR
Contributors	Antti Kallonen (TAU), Francesco De Pretis (TAU), Alpo Värri (TAU), Elina Nürenberg-Goloub (GUF)
Delivery date	31.7.2025
Dissemination level	Public
Type	Data sets

Version 01

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Revision history

Date	Authors	Revision
4.7.2025	Alpo Värri (TAU)	Draft version
17.7.2025	Antti Kallonen (TAU), Francesco De Pretis (TAU)	First version
22.7.2025	Elina Nürenberg-Coloub (GUF)	First revision
		Final version

Contents

Partner short names

GUF	Johann Wolfgang Goethe Universität Frankfurt am Main
accelCH	accelopment Schweiz AG
ESAIC	European Society of Anaesthesiology and Intensive Care
Fraunhofer	Fraunhofer Institute for Translational Medicine and Pharmacology ITMP
F4	F4 Pharma GmbH
TAU	Tampereen Korkeakoulusäätiö SR
UCD	University College Dublin
UMCG	Universitair Medisch Centrum Groningen
MiDA	Medical Intelligent Data Analytics GmbH
KC	Lietuvos Sveikatos Mokslu Universiteto Ligonine Kauno Klinikos
ICS-HUB	Hospital Universitari de Bellvitge
UMFCD	Universitatea de Medicina si Farmacie Carol Davila din Bucuresti
APHP	Assistance Publique – Hôpitaux de Paris

Abbreviations

D	Deliverable
LLM	Large language model

Executive Summary

The COVend project aims to demonstrate the efficacy of FX06 in treating Acute Respiratory Distress Syndrome (ARDS), initially targeting COVID-19 patients. However, as COVID-19 cases declined, the consortium made a strategic decision in 2023 to pivot the clinical trial from the COVID-19-specific IXION study to a mechanism-centred ARDS approach in IXION2.0. This pivot broadened patient eligibility to encompass ARDS of various aetiologies. Despite extending the recruitment period multiple times until December 2024 and conducting the trial across six sites in five European countries, enrolment remained critically low, with only 18 of the targeted 263 patients recruited (6.8%). The timeline constraints became particularly acute given that unblinding occurred in April 2025, with multiomics data only becoming available to WP6 in June 2025, leaving insufficient time for comprehensive data analysis within the project runtime. To address these challenges, TAU developed an innovative synthetic data approach, ensuring the project could still deliver valuable scientific outputs despite the limited real patient data.

This document describes the public release of synthetic datasets and related analysis results developed within the COVend project. These resources are specifically intended to support predictive modelling of ARDS mortality. The data and accompanying analyses, including logistic regression coefficients extracted using large language model (LLM) methodologies, have been made openly accessible via the Zenodo data portal.

1 Introduction

The primary aim of this deliverable is to ensure public and open access to the synthetic datasets and results from data analysis performed in support of the COVend Bayesian modeling activities. Specifically, these datasets and resources contribute directly to modeling ARDS mortality outcomes, advancing the goal of personalised medicine in treating ARDS.

1.1 Purpose and scope of the deliverable

The objective of this deliverable is to publicly share all datasets, scripts, and results necessary for replication, validation, and further research by the scientific community. By utilising Zenodo, the project ensures persistent and easily accessible resources for external stakeholders.

1.2 Context of COVend WP6 Deliverables

The COVend project has undertaken several key deliverables related to data and modelling, which belong to the context of this report:

D6.2 Research datasets from the models

This deliverable outlines the strategy for the publication of the IXION2.0 and synthetic data and models.

D6.3 Multivariate patient profile models for Decision Support in COVID-19 patients
D6.4 Time dynamics models for decision support

These deliverables provide detailed information on the modelling approaches and methodologies for decision-support

D6.7 Decision-support models validation report

This deliverable informs on validation results of the literature-based prior model and the posterior model refined using the IXION2.0 data.

D6.6 Web-based implementation of models for decision support

A web-based tool presented in this deliverable will exploit the datasets and models, enhancing accessibility and utility for researchers and clinicians.

D6.5 Model toolbox for the research community
D6.8 All relevant data and data-analysis results made available to a relevant data portal

These deliverables provide information on the publicly available models and data.

1.3 Outcomes

- Publication of synthetic datasets representing distinct clinical outcomes in ARDS patients under various therapeutic regimens.
- Logistic regression coefficients (priors) derived from published literature using a LLM.

- Metadata and Python scripts provided for reproducibility and transparency.
- Data and associated resources made openly accessible via Zenodo for long-term availability and citation.

2 Zenodo repository

All files described below are available for download at:

https://zenodo.org/records/15809173

2.1 Repository contents and file descriptions

The Zenodo repository contains the following items:

- **prior_coefficients_70b_llama.csv**
  A comma-separated values file listing the logistic regression coefficients (odds ratios and confidence intervals) extracted from medical literature by a 70-billion-parameter LLaMA model. These serve as prior parameters for Bayesian logistic regression analyses.
- **covend-data-generator-v3.py**
  A Python script that automates the generation of synthetic patient records. It ingests parameter definitions (from the accompanying JSON) and produces datasets simulating ARDS progression under various therapeutic scenarios.
- **parameter_info_v3.json**
  A JSON file defining all parameters used by the data generator, including variable names, distributions, ranges, and clinical interpretations. This metadata file ensures reproducibility and transparency in how synthetic values are sampled.
- **synthetic_data_effective.xlsx**
  An Excel workbook containing a synthetic cohort of patients simulated to represent effective FX06 therapy outcomes. Each row is a patient record with demographics, vital signs, and an outcome label indicating survival.
- **synthetic_data_ineffective_safe.xlsx**
  An Excel workbook with synthetic patient data modelling an ineffective yet clinically safe dosing regimen. It mirrors the same schema as the “effective” dataset but reflects parameter choices that do not improve outcomes.
- **synthetic_data_ineffective_unsafe.xlsx**
  An Excel workbook containing synthetic records for a regimen both ineffective and outside established safety bounds. This dataset illustrates risk scenarios and supports sensitivity analyses in the Bayesian framework.

3 Outlook

The openly published synthetic datasets and data generation pipeline represent a lasting contribution to ARDS research beyond the COVend project. These resources enable researchers to develop and validate new predictive models without privacy constraints, fostering innovation in algorithm development and educational applications. The modular Python pipeline can be adapted for other critical care conditions or enhanced with additional biomarkers as medical knowledge advances. As more researchers utilise and improve upon these tools, they will form the foundation for a growing ecosystem of synthetic medical datasets. This approach demonstrates how projects can navigate data privacy regulations while maintaining scientific transparency, potentially serving as a template for future EU projects facing similar enrolment challenges. The permanent availability through Zenodo ensures these resources remain accessible for advancing ARDS research and medical education after the project's conclusion.