

## SPECIAL REPORT

# Sharing is caring: how COVID-19 led to large-scale collaboration for icudata.nl

**P. Elbers, P. Thorai, T. Dam, L. Fleuren on behalf of the Dutch ICU Data Warehouse Collaborators**

Laboratory for Critical Care Computational Intelligence, Department of Intensive Care Medicine, Amsterdam Medical Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam and University of Amsterdam, Amsterdam, the Netherlands

## Correspondence

P. Elbers – p.elbers@amsterdamumc.nl

**Keywords** - data sharing, collaboration, machine learning

For many intensivists worldwide, the pandemic will have created long-lasting memories. Some of them grim, such as massively overwhelmed ICUs, the dehumanising appearance of seemingly interchangeable and mostly prone patients, and healthcare professionals hidden behind personal protective equipment. Some of them energising, including surprising patient recoveries as well as the unprecedented praise and recognition for intensivists and intensive care medicine from both the media and society at large.

For the Laboratory for Critical Care Computational Intelligence<sup>[1]</sup> at Amsterdam UMC, the pandemic proved to be nothing short of a rollercoaster ride. And despite the many challenges imposed upon our profession by the pandemic, this ride was largely fuelled by excitement, in particular the rapidly expanding enthusiasm for large-scale data sharing and collaboration between Dutch ICUs.

Obviously, our story started long before the COVID-19 crisis. Intensive care medicine is a natural habitat for data science as large amounts of data are routinely collected during intensive care treatment, such as those from devices for monitoring and life support. Our laboratory was created with the primary aim to unite clinical and data science expertise to use these data to improve the care and treatment of future critically ill patients. We do so by developing and validating models, integrating these into clinical decision support tools to be used at the bedside and evaluating their effect on outcomes relevant for critically ill patients.

Three of the most prominent results from our philosophy are AmsterdamUMCdb, the first freely available European ICU database under the European Society of Intensive Care Medicine / Society of Critical Care Medicine joint data sharing initiative,<sup>[2]</sup> bedside decision support for personalised antibiotic dosing<sup>[3,4]</sup> and bedside decision support for preventing untimely patient discharge from intensive care units (ICUs).<sup>[5]</sup>

Because of these contributions to the field, our lab had the infrastructure and knowledge base to readily facilitate large-scale data sharing when the pandemic hit the Netherlands. Specifically, our expertise could facilitate sharing of high-frequency device data and most other clinical information from the electronic health record (EHR), with the goal to generate insights from ICU patient data as the pandemic was unfolding. These data were thought to reflect the large variation in COVID-19 related clinical practice resulting from the limited and rapidly evolving COVID-19 evidence base and possibly the large variation in patient characteristics and outcomes between centres. These variations may be leveraged by advanced statistics and machine learning to determine optimal individual patient management.

Right from the start of the project<sup>[6]</sup> we experienced an unprecedented momentum. Data protection officers immediately offered help to provide a legal framework for responsible data sharing. Within days, our medical ethics committee approved our protocols. Data sharing agreements were drafted ensuring equal possibilities for data access for all participating ICUs. All hospitals in the Netherlands with an ICU were approached and documentation was reviewed locally before permission to participate was granted. With full support of the Dutch Society for Intensive Care (NVIC) and their research network RCCnet, 66 out of 81 ICUs confirmed their participation within weeks.

Template Structured Query Language (SQL) queries were developed to automatically extract EHR data for each of the major EHR systems used in the Netherlands: MetaVision, ChipSoft and Epic. Collected data cover the entire ICU stay and include demographics, data from devices for vital signs monitoring and life support, data on administered medication, laboratory results and data entered by the treatment team including clinical observations. All data were pseudonymised in the delivering hospitals.

An extraction, transformation and load process was designed to combine raw data from the different EHR systems. All parameters from the collaborating ICUs were manually reviewed and mapped

