

Category :[ICU organization](#)

A53 - Systematic review and comparison of icu datasets – a decision guide for clinicians and data scientists

CM Sauer¹; TA Dam¹; LA Celi²; AR Girbes¹; PW Elbers¹

¹*AmsterdamUMC, Department of Intensive Care Medicine, Amsterdam, Netherlands,* ²*Massachusetts Institute of Technology, Laboratory for Computational Physiology, Institute for Medical Engineering and Science, Cambridge, United States*

Introduction:

The publication of large, rich, single patient level data sets has significantly propelled clinical research possibilities in the intensive care unit (ICU) and beyond. After publication of MIMIC-III, additional datasets have previously been released, thus warranting a comparison of their data completeness and richness to allow scientists to choose the most appropriate dataset(s) for their clinical problem.

Methods:

A systematic search of published and pre-print articles was performed to identify all publicly available, adult, critical care, patient level databases. Subsequently, databases were compared using a priori defined categories, such as demographics, patient characteristics and data richness.

Results:

A total of 4 ICU databases were identified (MIMIC-IV, eICU-CRD, AmsterdamUMCdb, HiRID). Number of unique patient admissions varied significantly between datasets, with eICU-CRD being the largest (>130K) and AmsterdamUMCdb the smallest (>23k). ICU mortality and intensity of treatment also varied, with eICU-CRD 28-day mortality rates and frequency of ventilation being lowest among the datasets. Frequency of lab values tended to be highest in MIMIC-IV, while frequency of vital signs was highest in AmsterdamUMCdb.

Conclusion:

Several high-quality ICU databases are currently available. The research question, and thus required sample size, presence of covariates and frequency of measurements, should inform which database to use.

References:

- Johnson AEW, et al (2016) MIMIC-III, a freely accessible critical care database. *Sci Data* 3:1–9.
- Pollard TJ et al (2018) The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data* 5:1–13.
- Thoral PJ et al (2021) Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration. *Critical Care Medicine*
- Faltys, M., et al (2021). HiRID, a high time-resolution ICU dataset (version 1.1.1). *PhysioNet*