**MIMIC-II V3.0: An Open Dataset for Critical Care**

**Lu Shen\*, Mengling Feng\*, Mohammad Ghassemi\*, Roger Mark**

**\* These authors contribute equally to this work**

**Abstract**

*(170 words maximum, no references)*

Objective—We sought to develop an intensive care unit research database applying automated techniques to aggregate high-resolution diagnostic and therapeutic data from a large, diverse population of adult intensive care unit patients. This freely available database is intended to support epidemiologic research in critical care medicine and serve as a resource to evaluate new clinical decision support and monitoring algorithms.

Major Outcomes—The Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) V3.0 database consists of XXX intensive care unit stays. The investigators collected detailed information about intensive care unit patient stays, including laboratory data, therapeutic intervention profiles such as vasoactive medication drip rates and ventilator settings, nursing progress notes, discharge summaries, radiology reports, provider order entry data, International Classification of Diseases, 9th Revision codes, and, for a subset of patients, highresolution vital sign trends and waveforms. Data were automatically deidentified to comply with Health Insurance Portability and Accountability Act standards and integrated with relational database software to create electronic intensive care unit records for each patient stay. The data were made freely available in XXX through the Internet along with a detailed user’s guide and an assortment of data processing tools. The overall hospital mortality rate was XXX%,

which varied by critical care unit. The median intensive care unit length of stay was XXX days

(interquartile range, XXX-XXX days). According to the primary International Classification of

Diseases, 9th Revision codes, the following disease categories each comprised at least XXX% of the

case records: diseases of the circulatory system (XXX%); trauma (XXX%); diseases of the digestive

system (XXX%); pulmonary diseases (XXX%); infectious diseases (XXX%); and neoplasms (XXX%).

Conclusions—MIMIC-II V3.0 documents a diverse and very large population of intensive care unit

patient stays and contains comprehensive and detailed clinical data, including physiological

waveforms and minute-by-minute trends for a subset of records. It establishes a new public-access

resource for critical care research, supporting a diverse range of analytic studies spanning

epidemiology, clinical decision-rule development, and electronic tool development.

**Background & Summary**

*(700 words maximum)*

We report the establishment of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) research database that is notable for four factors: it is publicly and freely available to other research organizations upon request; it encompasses a diverse population of intensive care unit (ICU) patients; it contains high temporal resolution data, including laboratory results, electronic clinical documentation, and bedside monitor numeric trends and waveforms (such as the electrocardiogram); and it has been deidentified in a Health Insurance Portability and Accountability Act-compliant manner. The MIMIC-II database will support a diverse range of analytic studies spanning epidemiology, clinical decision-rule development, and electronic tool development. Historically, large-scale ICU databases have been effective resources to understand risk factors and natural histories of critical illness as well as the efficacy of various treatment strategies. For instance, Acute Physiology and Chronic Health Evaluation I–III and Project Impact contained daily abstractions of patient data that provided new insights and scoring tools to relate patient outcomes and lengths of stays with the patients’ conditions on admission (1, 2). Such collection and analysis of large volumes of ICU data are invaluable to the advancement of clinical knowledge, but it is extremely effort-intensive because there are substantial challenges to the collection of the data. Such difficulties include: disparate sources of data, eg, clinical documentation versus laboratory results; erroneous or missing data; unsynchronized time references; proprietary data formats; limitations of computing power, networking bandwidth, and digital storage capacity; and concerns related to patient privacy. The challenge of data collection has sometimes been addressed through coordinated efforts by a network of clinical investigators interested in specific problem domains such as acute respiratory distress syndrome (ARDSNET Trial) (3), acute kidney injury (4), or septic shock (5). However, these powerful disease specific databases were not designed to be exploited as research resources to support other domains of ICU research nor are their data widely available. In 2003, under National Institutes of Health funding, we established a research program with the objective of developing and evaluating advanced ICU monitoring and decision support systems. A critical requirement of our program was the development of a substantial and comprehensive clinical database from ICU patients. Now, 7 yrs later, the MIMIC-II database has reached a state of maturity sufficient to be made available to the wider research community. The database is intended to support a wide diversity of research in critical care. Unlike related databases, there are no access fees or extensive credentialing requirements, and documentation and other support are available so that the data will be accessible to the

largest community of researchers.

This article contains a detailed report of the MIMIC-II data acquisition process, which was

accomplished through collaboration among academic, industrial, and clinical groups.

Summary statistics are provided to characterize the database and we provide examples of

clinical hypotheses and physiologic signal processing algorithms we have studied with

MIMIC-II. The high temporal resolution parameters within the database such as hourly vital

sign trends, ventilator settings, intravenous medication drip rates, and fluid balances enable

novel investigations of transient clinical outcomes such as hypotensive episodes. Similarly,

MIMIC-II enables the analysis of transient independent variables such as electrocardiogram

waveform features and their associated clinical outcomes. The unique features of MIMIC-II

are compared with other major databases and we discuss the major challenges encountered

in developing MIMIC-II and explore future improvements. The MIMIC-II database takes

advantage of improvements in healthcare information technologies to establish a new

standard in public-access databases for critical care research.

**Methods**

Data collection

Data transformation & Merging

Data de-identification

**Data Records**

What data is in MIMIC-II V3.0?

Basic statistics

**Technical Validation**

Statistics comparing v3.0 to v2.6

Comparing our data to general knowledge

Repeat a previous study

**Usage Notes**

Data user agreement

Training requirement

**Acknowledgements**

**Author Contributions**

Each author’s contribution to the work should be described briefly, on a separate line, in the Author Contributions section: please see also the *Nature* journals' [authorship policies](http://www.nature.com/authors/policies/authorship.html).

'AK did this and that.'

'BG did this, that and the other.'

**Competing financial interests**

**Figures & Tables**