The eICU Collaborative Research Database, a freely available database for critical care research

Name Name[1]   Name Name[1]   Name Name[1]
Name Name[2]   Name Name[2]

April 24, 2017

Abstract

- Introduce the eICU Collaborative Research Database.
- Valuable dataset for research and education.

1 Introduction

Patients admitted to critical care are severely ill and require life-saving treatment. In 20??, critical care accounted for over 1% of the gross domestic product of the United States (US) (**???**). Intensive monitoring of patients is carried about by numerous equipment at their bedside, resulting in large quantities of data being collected continuously.

The Philips eICU telehealth program provides telemonitoring services to intensive care units (ICUs) across the US and internationally. The program allows caregivers from remote locations to monitor treatments for patients, alert caregivers to sudden deterioration, and provide recommendations regarding care plans.

During routine use of the eICU program, large amounts of data are collected and transferred to remote locations. This data is archived by Philips and

transformed into a research database by the eICU Research Institute (eRI). The laboratory of computational physiology (LCP) partnered with eRI to produce an open access database sourced from the eICU telehealth program.

- Previously shared MIMIC-III (Johnson et al. 2016) The shared database is referred to as the eICU Collaborative Research Database (eICU-CRD).

1.1 Patient characteristics

eICU-CRD comprises of 2??,??? patient encounters corresponding to 1??,??? unique patients. Patients were admitted to one of ??? hospitals located throughout the US. Table 1 provides demographics of the dataset, including hospital level characteristics.

Table 2 highlights the top 10 most frequent admission diagnoses in the dataset as coded by the APACHE IV diagnosis system (**???**).

1.2 Classes of data

Data includes vital signs, laboratory measurements, medications, APACHE components, care plan information, admission diagnosis, patient history, time-stamped diagnoses from a structured problem list, and similarly chosen treatments.

Periodic vital signs are recorded every 5 minutes and include heart rate, respiratory rate, oxygen saturation, temperature, invasive arterial blood pressure, pulmonary artery pressure, ST levels, and intracranial pressure (ICP). Periodic vital signs are not validated by care staff and not available for all patients. The most frequently available periodic vital sign is heart rate (??%), and the least available periodic vital sign is ICP (??%). Aperiodic vital signs are collected at various times and include non-invasive blood pressure, PAOP, cardiac output, cardiac input, SVR, SVRi, PVR, and PVRi. The most frequent aperiodic vital sign is mean blood pressure (??%), and the least frequent is PVRi (??%).

Laboratory values are collected within the hospital and are stored with the unit of measurement and last time the value was revised.

Treatments and diagnoses are selected by the care giver from a drop down list of options. There are ???? unique treatments and ???? unique diagnoses in the database. Both treatment and body system use a hierarchical coding system allowing for specification of terms corresponding to overarching groups such as body system. Care plans are similarly structured but do not have the same hierarchical grouping as treatments and diagnoses.

Table 2 gives overview of the classes of data available.

Data	Note
Admission diagnosis	
APACHE	
Care plans	
Diagnosis	
Lab	
Medication	
Past history	
Patient	
Treatment	
Aperiodic vitals	
Periodic vitals	

Figure 1 provides a visualization of data available for a single patient.

2 Methods

2.1 Database development

eICU-CRD was created according to a relational schema comprising of 17 tables. Figure ? shows the database schema.

2.2 Sample selection

A representative sample of patients was extracted from the eRI data warehouse. This sample of patients was selected by …

2.3 Deidentification

All tables were de-identified to meet the safe harbor provision of HIPAA. These provisions include the removal of all personally identifiable patient identifiers, ages over 89, and other personal information. Data was de-identified using a python package based upon a publicly available de-identification package which has previously been used to de-identify electronic health records (**???**). Large portions of all tables were manually reviewed by at least three personnel to verify all data had been de-identified.

2.4 Code availability

As the aim of the eICU-CRD is to foster collaboration in secondary analysis of electronic health records, we have created an openly available repository for sharing code (**???**). Code used to extract concepts from the data is under continuous development.

3 Data records

Each patient stay in a unit, where the primary unit of care is the ICU, is identified by a single integer: the `patientunitstayid`. Each unique hospitalization is also assigned a unique integer, known as the `patienthealthsystemstayid`. Each `patienthealthsystemstayid` has at least one or more `patientunitstayid`. Finally, patients are identified by a `uniquepid`, and each `uniquepid` can have multiple hospital and/or unit stays. All tables use `patientunitstayid` to identify an individual unit stay, and the patient table can be used to determine unit stays linked to the same patient and/or hospitalization.

4 Technical validation

- Changes were kept to a minimum
- Issue tracking used to keep track of issues
- Compared distribution of X to MIMIC. Figure shows…

5 Usage notes

5.1 Data access

Data can be accessed via a PhysioNetWorks repository (**???**). Usage of the data requires proof of completion for a course on using human subjects in research. Usage of the data also requires signing of a data use agreement stipulating, among other items, that the user will not share the data and will not attempt to re-identify any patients or institutions.

5.2 Example usage

- Examples of usage include…

5.3 Conclusion

- Shared a unique resource
- Move towards end-to-end reproducible research

Acknowledgments

The authors would like to thank…

Funding

Author contributions

XXXX

Competing interests

References

Johnson, A.E. et al., 2016. MIMIC-iii, a freely accessible critical care database. *Scientific data*, 3.