MIMIC code: a repository for deriving clinical concepts

Alistair E. W. Johnson [a], Tom J. Pollard [a]

[a] *Massachusetts institute of Technology, Cambridge*

Abstract

Secondary analysis of routinely collected data - contrasted with the primary analysis conducted in the process of caring for the individual patient - offers an opportunity to extract knowledge that will lead us towards the goal of optimal care. Retrospective studies frequently require similar clinical concepts, so there is benefit in providing open, standardized tools for deriving these concepts to ensure the consistency and efficiency of future studies. We present the MIMIC Code Repository, a collection of open source code for deriving clinical concepts using the MIMIC Critical Care Database. Concepts include severity of illness scores, organ failure indices, and duration of treatments such as ventilation and dialysis. The MIMIC Code Repository is in active development on GitHub. All code is made available under a permissive MIT license unless otherwise indicated.

Key words: critical care; open data; data mining; secondary use of electronic health records.

1 Introduction

Intensive care units care for the sickest patients in a hospital. Often require support for multiple organs. High mortality rate. High rate of sepsis. Potential for improving outcomes...

There is substantial heterogeneity in intensive care populations, particularly in aspects such as patient physiology, presence of disease, and intervention types. This heterogeneity presents a significant challenge to understanding the relationships between provision of care and patient outcomes, and as a result the impact of many widely-practiced treatments and interventions remains

Email addresses: `aewj@mit.edu` (Alistair E. W. Johnson), `tpollard@mit.edu` (Tom J. Pollard).

unknown [REFS: Ioannadis]. It has been estimated that [X in X treatments are not supported by evidence].

Vast quantities of data are routinely collected by modern hospital monitoring systems and even more so in intensive care units where patients frequently suffer organ failure and require close observation. There is optimism that increasing availability of large scale clinical databases will offer opportunities to overcome many of the challenges associated with heterogeneity and offer new insights into critical care medicine [REF: BIG DATA etc].

But in addition to quality data, robust code must be developed to enable analysis. Challenges in interpreting data include deriving severity of illness scores (often used in adjustment), deriving duration of treatments based on [ambiguous?] data, and selecting cohorts based on clinical guidelines. The code used to conduct analysis is critically important in the results.

Typically concepts are coded during analysis, often independently by research groups, often without complete understanding of the data and unable to interact with hospital staff. As a result, the quality of code varies widely. [For example, of X papers using MIMIC in 2013, X different approaches to identifying sepsis were taken.] While controversial, this point was highlighted by NEJM. Numerous examples where errors in code have resulted in retracted papers...

We have implemented an open source library. Novel because... Able to work directly with clinicians to develop the code. Bridge gap between data scientists and clinical environment. Support interpretation of data. Ref NEJM paper highlighting importance of working with clinicians. Also bridging the gap paper. Novel because entire workflow can be reproduced, from data to publication.

While the case for open data is well publicised, we believe open code is equally important.

2 Results

A prerequisite for using much of the code in the MIMIC Code Repository is access to the MIMIC-III Database, so we provide scripts to enable researchers to build the MIMIC database in a variety of database systems including PostgreSQL, MySQL, Oracle, and MonetDB. The set of core clinical concepts which have been extracted using structured query language (SQL) [1] are as follows. Additionally, we seek to provide introduction to the data. Tutorials

[1] All queries have been developed and tested using PostgreSQL 9.5.1.

are provided to give new users and introduction to the MIMIC database. A cookbook of sample code is provided to introduce new users to the database in a friendly way.

2.1 Severity of illness scores

Several scoring systems have been developed in order to quantify extent of illness in hospital patients [REF]. These scoring systems are used widely in secondary analysis of health data for a variety of purposes, including selection of study populations and as covariates for severity adjustment in physiological models. While severity scores are integral to many research studies, their definition can present several challenges and it is crucial to recognise the limitations of how scores are generated. Firstly, for example, most severity scores are developed with well curated datasets, usually acquired either through prospective data collection by trained personnel or through manual data abstraction by qualified professionals. As a result, the data tends to be cleaner and often has, perhaps more importantly, a distribution that is markedly different from routinely collected data such as that present in an electronic health record.

Secondly, routinely collected data often lacks some of data elements required to compute the score. For example, the comorbidity “”biopsy proven cirrhosis“” is not simple to determine as there is no routine documentation of this concept in the clinical workflow. Finally, the data definitions for the same concept can vary between the original dataset used to define the severity score and the electronic health records being analyzed. For example, the Glasgow Coma Scale (GCS), a common marker of neurological dysfunction which ranges from 3 (worst) to 15 (best), is usually assumed to be 15 for patients who are unable to be assessed due to sedation or ventilation. In an electronic health record however, this definition is not strictly adhered to as there is no defined protocol, and as a result sedated patients may be assigned a score of 15 by some care providers, and a score of 3 by others.

Working directly with caregivers has helped us to address these issues in the code, helping to ensure the derived severity scores accurately reflect the true severity of illness in patients. There are five severity of illness scores currently implemented in the MIMIC Code Repository: APS-III [**?**], SAPS [**?**], SAPS-II [**?**], SOFA [**?**] and OASIS **??**. A more detailed comparison of the severity scores is provided in the supplementary material, along with discussion of the assumptions that have been made when defining severity scores. An example of the importance of the variance caused by a non-centralized code base is shown in **??**, where the performances of two different implementations of the SOFA score in discriminating hospital mortality are shown. Both of these implementations have been used in previous publications.

2.2 Duration of interventions

When carrying out a study it may be necessary or desirable to know the durations of interventions and physiologic states such as mechanical ventilation, vasopressor administration, and hypotension. Deriving these durations from routinely collected data is often non-trivial, requiring a strong understanding of the underlying data as well as the environment in which it was collected. Cao et al highlight this issue in a 2010 paper on heuristics to determine ventilation times of ICU patients, noting that while “”On the surface, it appears trivial to determine ventilation times“”, “”when facing the reality of retrospective data, it is not straightforward“” [REF Cao et al 2010]. In collaboration with staff at the Beth Israel Deaconess Medical Center - the primary source of data in the MIMIC-III database - we have provided code to derive start and stop times for continuous events such as mechanical ventilation, dialysis, and various vasopressors.

Taking invasive mechanical ventilation as an example, intubation and extubation times are not well documented, so they are determined through the use of surrogate settings which are synchronized with the patient's chart via the ventilator. These settings include flow rates, tidal volumes, and so on. The key assumptions made when defining ventilation are: start time of ventilation is the first occurrence of a ventilator setting, and end time of ventilation is the time of a ventilator setting, followed by at least 8 hours of no ventilator settings. Consequently, short extubations (¡8 hours) are not captured by the ventilation query, and the end time of ventilation only approximates the actual extubation time. An example of these durations is provided in Figure **??**.

2.3 Clinical guidelines and definitions of disease

Numerous clinical guidelines have been developed by well-recognised experts and organisations to assist in the identification and management of specific clinical conditions. These guidelines are often used in clinical studies for risk adjustment and for selection of patient cohorts. One clinical definition implemented in the MIMIC Code Repository, for example, is Angus criteria. These criteria are a widely used definition of severe sepsis, a high-risk complication of infection that consumes considerable healthcare resources and is strongly associated with patient mortality, with a 2001 paper attributing 215,000 deaths from severe sepsis in the US annually [REF - Angus 2001]. The Angus criteria are based upon hospital billing codes, making them relatively simple to implement, but other guidelines are less straightforward.

The Glasgow Coma Scale (GCS), for example, represents the level of conciousness of a patient, and as such it is influenced by level of sedation. In the collection of data for severity scoring, values of GCS were set to 15 (normal) if the care provider felt the GCS was not a true reflection of the patient's neurological status. This situation would occur if the patient was sedated or if a tracheostomy prevented a verbal response. However, these values in MIMIC-III are often recorded as 3 (extremely abnormal) - in particular, the string for verbal response can be either “”1.0 No response“” or “”1.0 ET/Trachy“”. Simply extracting the GCS as it appears in MIMIC would be incongruent with the original definition of the scale and would likely compromise their discrimination and calibration. Ideally, all patients who have low GCS due to sedation would have their value replaced by 15, but in practice determining sedation status from a patient's chart is a difficult task.

In addition to implementing code for the Angus criteria for severe sepsis and Glasgow Coma Scale for neurological status, we provide scripts for a growing number of additional clinical guidelines. These include the Kidney Disease: Improving Global Outcomes (KDIGO) classification for acute kidney injury, a common, harmful, and potentially treatable condition characterised by abrupt decrease in kidney function [REF - KDIGO guidelines], as well as the Model For End-Stage Liver Disease (MELD) Score, which is commonly used in the care of patients with cirrhosis for assessing the severity of chronic liver disease. Critically ill patients frequently have many comorbidities which influence both their overall health and their trajectory of health during an individual hospital stay. To support analysis that seeks to capture the variation in patient comorbidities, the MIMIC Code Repository includes code for computing the Elixhauser Comorbidity Index, a clinical definition that seeks to summarize the level of comorbidities in individual patients using billing codes collected at hospital discharge [REFS].

3 Discussion

4 Materials and methods

Acknowledgements

The authors would like to thank Professor Roger G. Mark, the MIT Laboratory for Computational Physiology, Philips Healthcare and the Beth Israel Deaconess Medical Center for the creation of the MIMIC-III database.

Funding

This work has been supported by grants NIH-R01-EB017205, NIH-R01-EB001659, and NIH-R01-GW104987 from the National Institutes of Health.