

# The MIMIC Code Repository: Enabling reproducibility in critical care

Alistair E. W. Johnson<sup>1\*</sup>     David J. Stone<sup>2</sup>     Leo Anthony Celi<sup>1,3</sup>  
Tom J. Pollard<sup>1</sup>

<sup>1</sup> Massachusetts Institute of Technology, Cambridge, USA

<sup>2</sup> University of Virginia School of Medicine, Charlottesville, USA

<sup>3</sup> Beth Israel Deaconess Medical Center, Boston, USA

\* Corresponding author: `aewj [at] mit [dot] edu`

## Abstract

*Objective:* Lack of reproducibility in medical studies is a barrier to the generation of a robust knowledge base to support clinical decision-making. In this paper we outline the MIMIC Code Repository, a centralized code base for generating reproducible studies on an openly available critical care dataset.

*Materials and Methods:* Code is provided to load the data into a relational structure, define useful extractions of the data, and reproduce entire analysis plans including research studies.

*Results:* Concepts extracted include severity of illness scores, comorbid status, administrative definitions of sepsis, physiologic criteria for sepsis, organ failure scores, treatment administration, and more. Executable documents are used for tutorials and also reproduce published studies end-to-end, providing a template for future researchers to replicate. The repository's issue tracker enables community discussion about the data and concepts, allowing for users to collaboratively improve the resource.

*Discussion:* The centralized repository provides a platform for users of the data to interact directly with the data generators and facilitating greater understanding of the data. It also provides a location for the community to collaborate on necessary concepts for research progress and share them with a larger audience. Consistent application of the same code for underlying concepts is a key step in ensuring research studies on the MIMIC database are comparable and reproducible.

*Conclusion:* By providing open source code alongside the freely accessible MIMIC-III database, we enable end-to-end reproducible analysis of electronic health records.

# 1 Introduction

Concerns about the reproducibility of results in science are becoming increasingly prominent in both scientific and mainstream literature (1). Some commentators have gone so far as to call the current state a ‘crisis’, citing causes such as pressure to publish positive results, the cost of replicating studies such as double blind randomized controlled clinical trials, and the lack of emphasis on reproducibility as a requirement for sound science.

In parallel, health care has been undergoing a digital revolution in recent years. The Health Information Technology for Economic and Clinical Health Act has catalyzed the transition of hospitals and care institutions from paper based systems to electronic ones (2). Vast quantities of digital data are now routinely collected by modern hospital monitoring systems, even more so in intensive care units (ICUs) where patients 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 lack of evidence in medical practice.

The Medical Information Mart for Intensive Care (MIMIC-III) database is an example of such a data repository (3). The database comprises detailed clinical information regarding over 60,000 stays in intensive care units (ICU) at the Beth Israel Deaconess Medical Center in Boston, MA, USA, collected as part of routine clinical care. The MIMIC-III dataset is freely available to researchers around the world and has been widely used in the development of predictive models, epidemiological studies, and educational courses.

Perhaps the most important insight since the database was made open-access is how challenging research using EHR can be, requiring close collaboration between domain experts and data scientists. As MIMIC-III is a de-identified version of raw data stored during routine clinical care, there is a non-trivial body of work required to transform the data into a usable form for research. This derivation of clinical concepts on an EHR database is a resource-intensive task, however, and is a significant barrier to those unfamiliar with the clinical environment or the database structure. Moreover, if concepts are not defined collaboratively with those who are familiar with the workflows, including how the data is captured, the validity of the findings may be suspect.

In this paper, we describe the MIMIC code repository, a centralized location for derived concepts that are relevant to critical care research. Detailed descriptions on how the concepts are defined and extracted from the database are provided, including the assumptions that are made and the conditions for which a code or query is valid. Additional tools are provided to educate researchers on best practices for conducting a fully reproducible study using the database. The code is open source, follows good documentation practices, and is contributed to by members of the research community using MIMIC-III.

The repository provides a framework for collaboration around research. While the case for open data has been already been strongly made elsewhere, we believe *open code* is equally important. We would make the argument that the use of an openly available code repository will improve secondary analysis of health data by accelerating the understanding of datasets by researchers, and improving the consistency and validity of future studies.

## 2 The MIMIC Code Repository

The MIMIC code repository is available online (4) and is open source. Code is available as standardized scripts in languages including SQL, Python, and R. Scripts are modified to allow an individual who has been granted access to the MIMIC-III database to generate a number of “views” of the data, with each view being an extraction from the raw data. Each script is associated with an automatically generated unique commit hash that acts as an identifier for the code. Publications that use the code repository can further cite the commit hash, allowing other researchers to download a copy of the code used regardless of any modifications since. All code follows the principles of good scientific programming as outlined by Wilson et al. (5), including incremental development with a distributed version control system, unit tests, and a public issue tracker. The repository was tested on MIMIC-III v1.4 at the time of this publication.

There are three components to the repository that facilitate navigation of the data for research purposes. These components are:

- *Concepts*: code to extract important concepts from the health records. For example, a module on acute kidney injury (AKI) uses the criteria as specified by the Kidney Disease Improving Global Outcomes (KDIGO) and provides the code to identify patients with AKI in MIMIC.
- *Executable documents*: notebooks that allow text and analytical code to be seamlessly combined into a single executable document, allowing studies and tutorials to be reproduced.
- *Community*: public discussions to facilitate contributions from members of the MIMIC research community

### 2.1 Concepts

Code to extract concepts that are broadly applicable to research questions in critical care are provided in the repository. For example, severity of illness scores are frequently required to adjust for confounding factors in a study, but are complex to derive, and so scripts are provided for reuse. These and other concepts are coded in a modular fashion to reduce redundancy in code and allow for extension. The following sections describe various concepts currently available in the repository.

#### 2.1.1 Severity of illness scores

Severity of illness scores have been developed over recent decades to provide an assessment of the patient’s acuity, particularly but not exclusively, at the time of admission to the ICU (6). The principal aim of these scores is for risk-adjusting patient populations for benchmarking and research purposes such as comparison of cohorts in clinical trials and observational studies. In the context of performing research using MIMIC-III, the use of severity of illness scores for risk-adjustment is almost always required to address confounding.

While severity of illness scores are integral to risk adjustment, their calculation, if done retrospectively, presents challenges. Most severity scores were developed with well-curated datasets, put together through prospective data collection or manual data abstraction by dedicated trained personnel. 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 data elements required to compute the score. For example, the comorbidity “biopsy proven cirrhosis” is required for the Acute Physiology Score and Chronic Health Evaluation (APACHE) system, but this concept is not documented in a structured manner during routine care. 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. To illustrate this potential disparity, 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, but otherwise appear to be neurologically intact. 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 with local nurses and doctors has helped us to address these kinds of issues that potentially impact the code, helping to ensure the derived scores accurately reflect the true severity of patient illness. There are five severity of illness scores currently implemented in the MIMIC Code Repository: APS-III (7), SAPS (8), SAPS-II (9), and OASIS (10). A more detailed comparison of the severity scores is provided in the supplementary material, along with discussion of the assumptions made in calculating the scores. Organ dysfunction scores are also available and detailed later.

Each score is comprised of at least ten independent components. The APS III, SAPS II, SOFA, LODS, and OASIS scores are generally calculated using data from the first 24 hours of the patient’s stay. SIRS and qSOFA are screening tools with scores calculated on admission to the ICU which is concretely defined as up to 2 hours after the admission time. The distribution of these scores is shown in in Figure 1.

### **2.1.2 Organ dysfunction scores**

Organ failure is a hallmark of acute illness and is quantified in numerous scores. Some scores assess multiple organ systems: the Sequential Organ Failure Assessment (SOFA) score (11) and Logistic Organ Dysfunction System (LODS) (12) both assess six organ systems for failure. Others are organ specific. Examples include the Model for End-stage Liver Disease (MELD) (13), the Risk/Injury/Failure/Loss/End stage renal disease (RIFLE) criteria (14), the Acute Kidney Injury Network (AKIN) classification (16), and the Kidney Disease Improving Global Outcomes (KDIGO) criteria (17). The latter three scores assess the degree of acute kidney injury in a patient. A variety of lab, diagnostic, and therapeutic data are needed to calculate these scores.

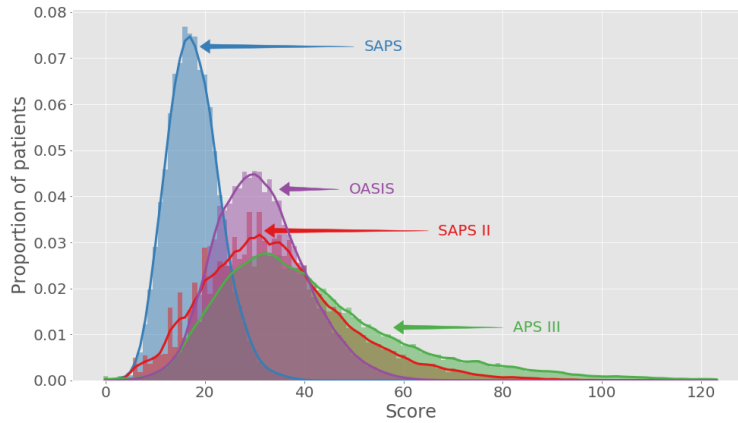


Figure 1: Comparison of severity of illness score distributions.

To highlight the discrepancies that can arise from the way a concept is defined, we contrast two versions of the SOFA score: one derived by prior researchers, and one available in the MIMIC code repository. Figure 2 shows the area under the receiver operator characteristic curve (AUROC) for hospital mortality for patients admitted in the MIMIC-III database between 2001-2008 using two versions of SOFA, grouped by the year of admission.

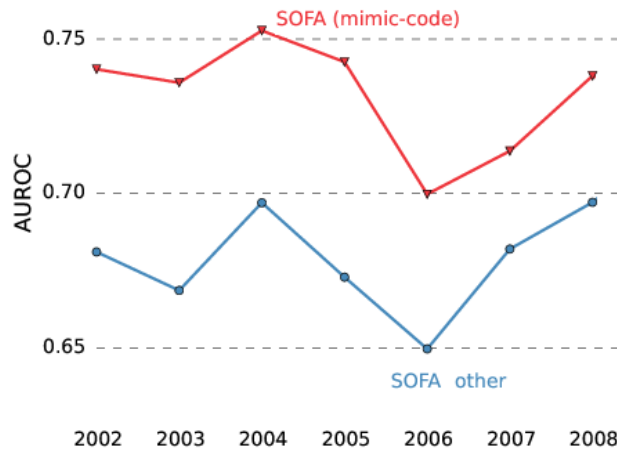


Figure 2: Comparison of AUROCs for SOFA scores calculated from mimic-code and a prior research report.

The disagreement between the two modalities is multifactorial, but a major contributing factor relates to an important variable: the Glasgow Coma Scale (GCS). In the original paper describing the SOFA score, clinicians were instructed to set GCS to its maximum value (15) if they were unable to assess the patient fully (for example, as a result of sedation to facilitate mechanical ventilation). In contrast, the documentation of GCS for these patients

in the MIMIC-III database is usually a value of 3, the minimum value, with a note that they are unable to assess the patient. Naive use of GCS values results in a dramatic differences in the capability of the score to discriminate severely ill patients and highlights the need to understand variables and how they are captured or derived. In the MIMIC Code Repository, special extraction steps are used to detect a GCS value of 3 due to sedation, and these values are corrected to 15 in the calculation of scores.

### 2.1.3 Timing of treatment

The timing and duration of treatment is an important concept for researchers seeking to understand issues that relate to the intensity of an administered intervention. Duration may serve as an indirect metric of severity and has been used in the development of decision support tools (18).

As a result of data capture limitations in the hospital, exact timing and duration of many medications and treatments are not explicitly available and so must be derived. Derivation may involve identification of surrogate data, known to be carried out with a high level of compliance, documented by clinical staff contemporaneous to the treatment. Figure 3 shows a schema for the derivation of the start and stop times of mechanical ventilation. Similar rules are used to define the timing of vasopressor administration and CRRT available in the repository. Clinical expertise is invaluable in developing these rules and interpreting the fine points of the medical chart that determine them.

An example of a patient undergoing mechanical ventilation and receiving vasopressor agents is provided in Figure 4.

### 2.1.4 Sepsis

Sepsis is a major and costly disease in the ICU, costing over \$20 billion in the US in 2011 (5.2% of all US hospital costs) (19), and growing to over \$23 billion in 2013 (6.2% of all US hospital costs) (20). Sepsis has traditionally been defined as the concurrent presence of systemic inflammation and infection, but a recent re-examination of the problem has suggested redefining the disease as a life-threatening organ dysfunction caused by a dysregulated host response to infection (21). The precise onset of sepsis is not typically documented in the EHR, and is, in fact, a difficult item to capture clinically. In their quantitative evaluation of septic patients, Seymour et al. (22) first identified patients suspected of infection by cross-referencing antibiotic use with requesting a microbiology assessment. We have implemented a similar approach, defining suspected infection as the acquisition of a microbiology culture followed by or shortly after ICU admission. Using this definition, and following the Sepsis-3 guidelines, we define sepsis as suspicion of infection associated with organ failure as quantified by an increase in SOFA  $\geq 2$ . This definition is admittedly a proxy for the actual onset of sepsis, but in the absence of more precise markers it serves as an approximation of onset time and could be used for development of decision support tools. Scripts for these concepts are available and a notebook describing the derivation is also available.

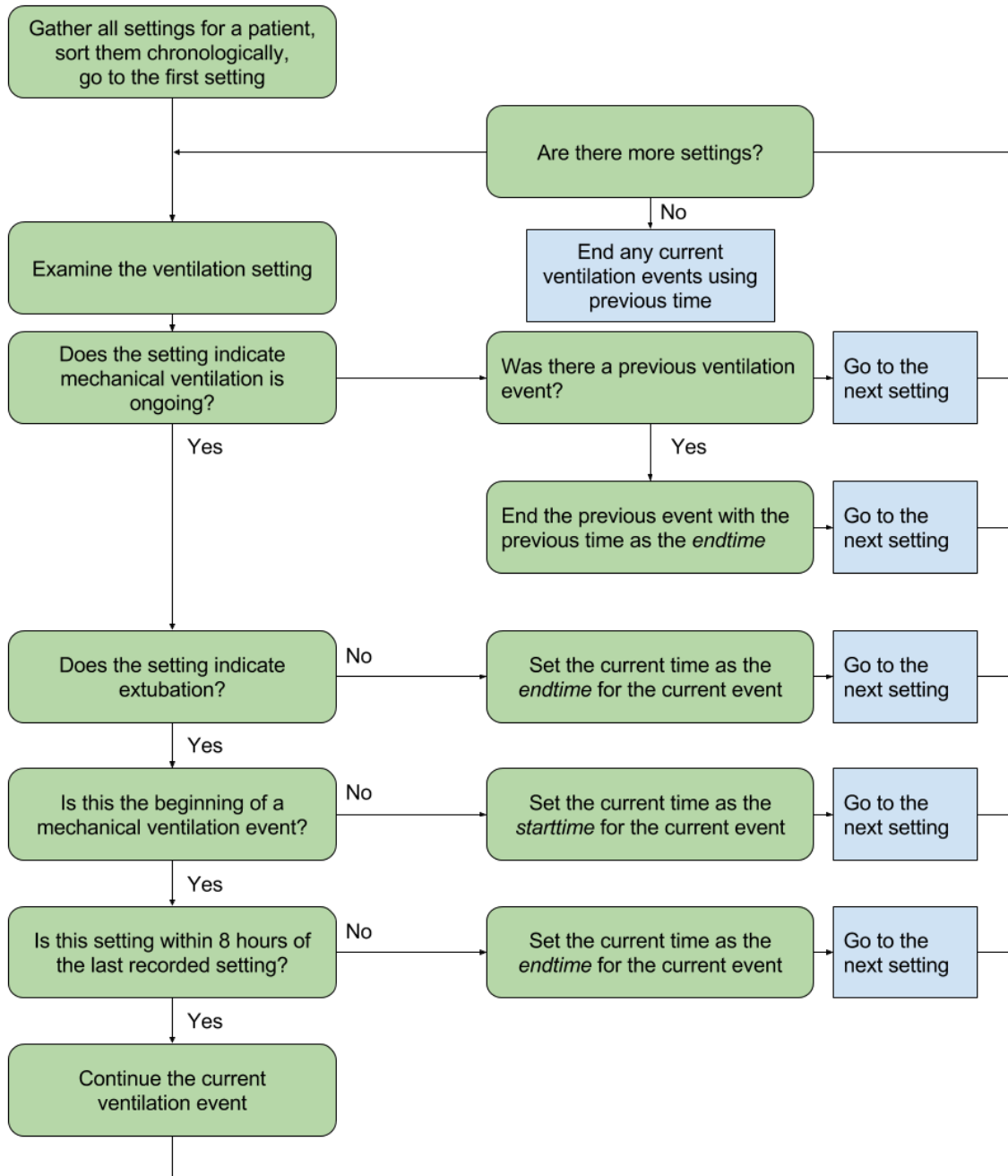


Figure 3: Logic behind the query for converting aperiodically recorded ventilator settings into durations of mechanical ventilation.

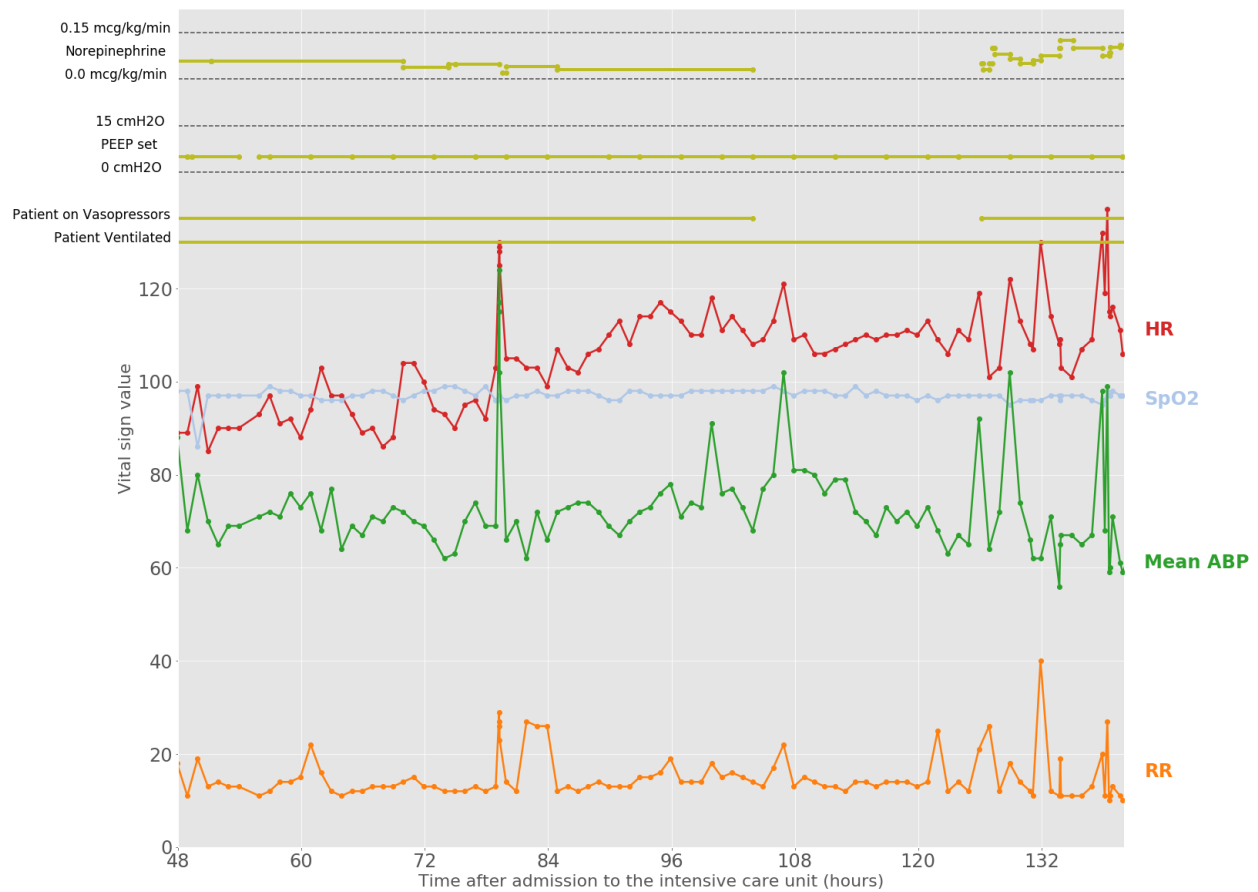


Figure 4: Example of a patient who was both mechanically ventilated and receiving vaso-pressors for cardiovascular support.



Identification of sepsis has also been done retrospectively using administrative data, and in particular billing codes acquired on hospital discharge. Two billing codes explicitly denote sepsis (ICD-9 codes 785.52 and 995.92). Angus et al. (23) and Martin et al. (24) describe algorithms for defining sepsis using a set of diagnostic and procedural ICD-9 codes. The criteria as proposed by Angus et al. (23) were validated in a later study by Iwashyna et al. (25). These three criteria; explicit coding, those as proposed by Angus et al. (23), and those proposed by Martin et al. (24), are available in the repository.

### 2.1.5 Comorbidities

Many ICU patients have chronic conditions prior to their acute presentation that affect their probability of surviving critical illness. Elixhauser et al. (26) codified these comorbidities into 29 categories using administrative data, specifically ICD-9 codes. The American Health and Research Quality group (AHRQ) continues to maintain these administrative codes via the Healthcare Cost and Utilization Project (HCUP), adapting them accordingly as changes are made to diagnosis and treatment coding (27). Finally, Quan et al. (28) proposed an enhanced ICD-9 coding methodology based upon examining inconsistencies among previous definitions. Diagnosis related groups (DRG), which are used to bill for the principle diagnosis for a patient hospitalization, are used to filter out those conditions that are not present prior to hospitalization. A comparison of these three methods is provided in Figure 5. These representations of comorbidities are provided in the repository, both with and without DRG filtering.

Van Walraven et al. (29) later aggregated comorbidities codified by Elixhauser et al. (26) into a single point score for in-hospital mortality prediction which is also available in the repository.

## 2.2 Executable documents

When both data and code is freely available to researchers - as is now the case for MIMIC-III - this provides a framework that allows a study to be entirely reproduced. This is especially powerful when toolkits such as R Markdown and Jupyter Notebook are employed, allowing documentation and code to be seamlessly combined to create executable documents. Figure 6 shows an example of a Jupyter Notebook that extracts patient demographics and displays the results for the user to view. Jupyter Notebooks are language agnostic, supporting code written in Python, R, MATLAB, SAS, and others (30).

We have found executable documents particularly valuable for research in cross-disciplinary fields such as healthcare, because they facilitate collaboration between data analysts and domain experts. Notebooks primarily serve three purposes: (i) they allow documentation of the logic behind the code in an organized and easy to read manner; (ii) they aid rapid writing of the code particularly during group discussions; and (iii) they provide a means of sharing details of a published study that captures the learning that takes place during the evolution of a research project. To encourage sharing of research code, we have reproduced a previously published study on indwelling arterial catheters and their association with in-hospital

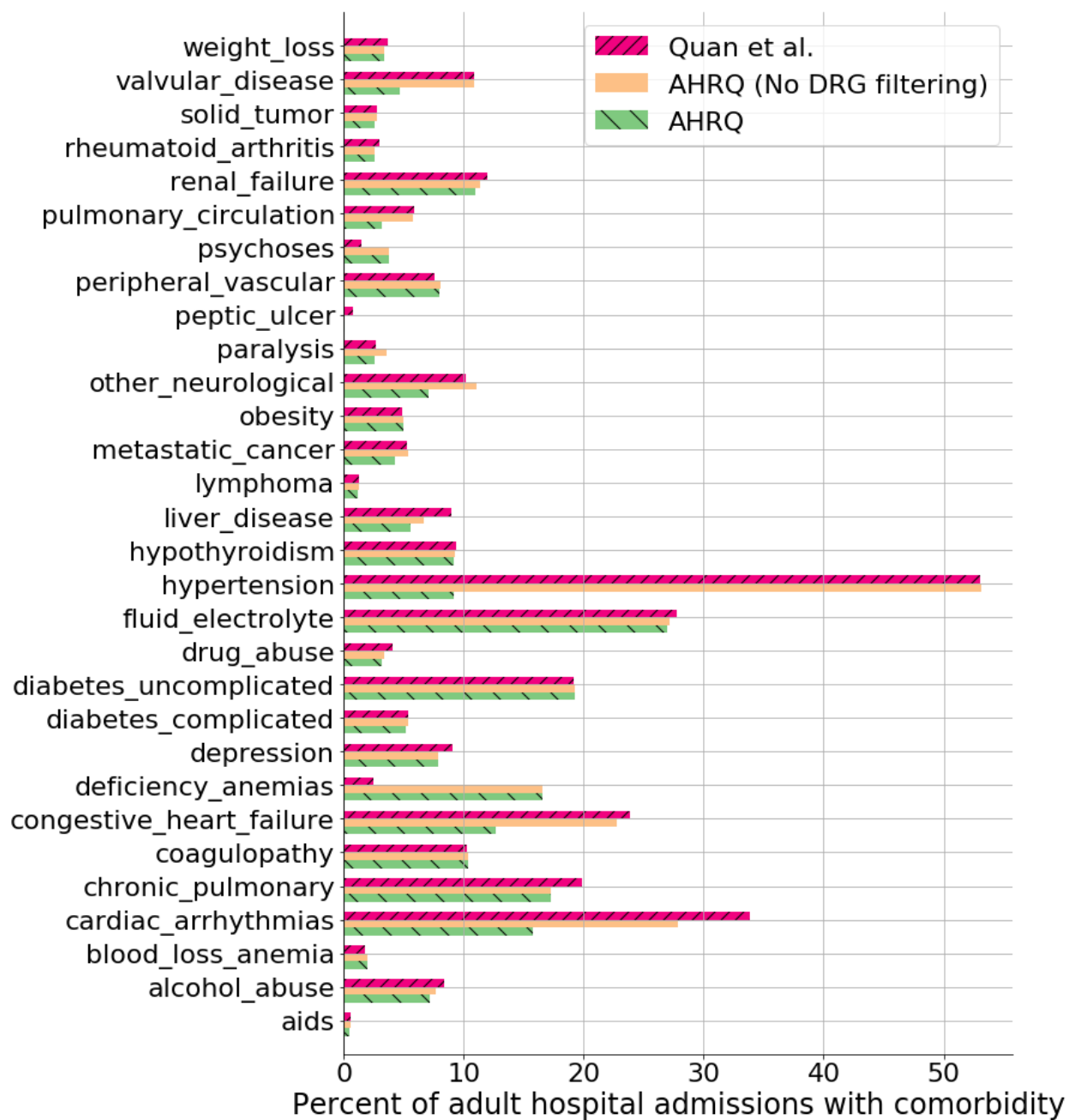


Figure 5: Comparison of three methods for calculating presence of a comorbidity for a patient using billing data: an updated coding from the AHRQ which uses DRG codes to mask non-comorbid conditions, the same coding without the DRG masking, and finally an alternative coding which does not use DRG masking proposed by Quan et al. (28).

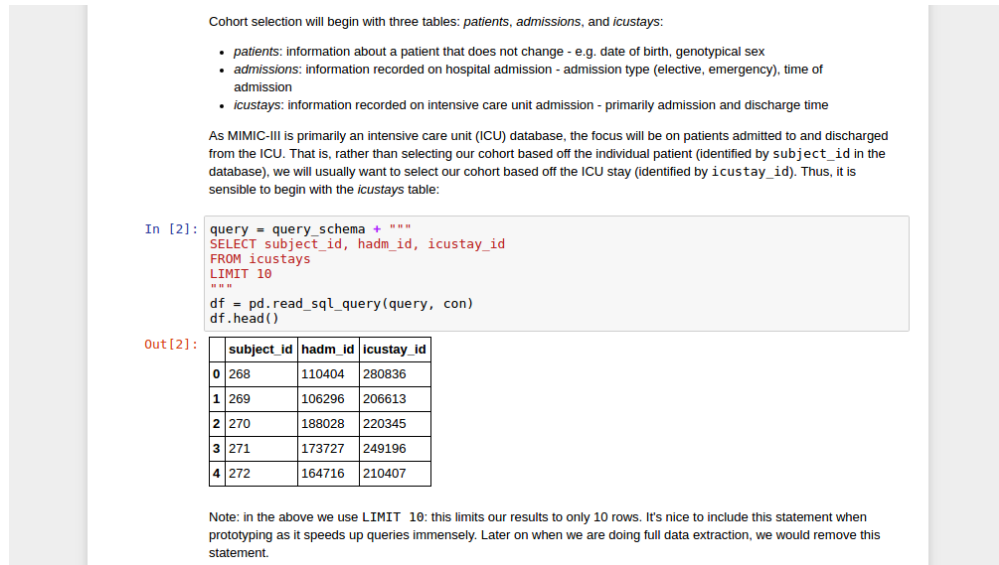


Figure 6: Example of a notebook providing a tutorial with MIMIC-III data.

mortality for haemodynamically stable patients with respiratory failure (32). This study was initially performed in MIMIC-II, which has since been superseded by MIMIC-III. As the structure of the databases differ, the study was reimplemented based on the manuscript. The executable documents perform the data extraction, preprocessing of the data, construction of a propensity score, and provide interpretation of the results. Specifically a Jupyter Notebook, `aline.ipynb`, extracts the study population and necessary data, outputting data to a plain-text file. An R Markdown file, `aline.Rmd`, subsequently loads data from the plain text file and tests the study hypothesis after matching cohorts with a propensity score. The executable documents provide a template for creating completely reproducible studies using the MIMIC-III database.

Executable documents are also an platform well-suited to tutorials. Harmonization of text and code allows for explanations of the subject matter, while the interactive nature of the document allows for experimentation and facilitates learning. A number of tutorials have been made available to explain key concepts important for working with MIMIC. For example, the transformation of recorded clinical parameters, such as hemofiltration settings, into desired clinical concepts, such as length of continuous renal replacement therapy (CRRT), is non-trivial and requires both domain and database expertise. An executable document is provided which overviews the process of exploring MIMIC-III, assessing the data stored within and creating the definition of CRRT provided in the database. In addition to explaining the logic behind the definition of CRRT, the tutorial also acts as a template for defining other concepts in the MIMIC database and potentially other similar ICU EHRs. Other tutorials include an introduction to Structured Query Language; a step by step guide to selecting a study cohort, and an outline of the data capture process for commonly recorded parameters in the database.

## 2.3 Community

The MIMIC code repository provides many previously mentioned benefits regarding distribution of the source code and enhancing reproducibility. An additional advantage is the communication channel opened between the maintainers and distributors of the MIMIC database and the users. Longo et al. (33) argue, in a well publicized editorial on data sharing, that researchers not involved in the collection of data may lack understanding of its underlying detail. Our framework connects researchers who reuse the MIMIC-III dataset with the laboratory and clinical staff who collect and produce the data, helping to provide context for downstream data analysis. Researchers can post issues inquiring about aspects of the data collection and best practices for analyzing the data and experienced users, some of whom involved in the collection of the data, can provide insight and advice. This correspondence facilitates appropriate and meaningful use of the data, and as all discussions are publicly available, the result in an organically growing set of documentation which spans both narrow and broad topics. Contributions to the MIMIC code repository by researchers are encouraged, progressively improving the codebase and helping to accelerate research in critical care. Source code control allows for transparency both in the authorship of the code and in the nature of any changes (34)

## 3 Conclusion

Transparent research processes can help to improve the quality of evidence that underpins health care, and the case for open data has been quite well described (35). To achieve full transparency, researchers must be able to provide both the data used for analysis *and* the code used to process it. The MIMIC-III database is exceptional due in no small part to its publicly accessible nature: all researchers who undergo human subjects research training and who sign a data use agreement can freely access the data. By supplementing the MIMIC-III Database with the MIMIC Code Repository, we provide a framework to allow completely reproducible research in critical care. Examples of reproducible code and even reproducible studies are available and provide a framework for future work with the database. While cultural barriers exist which may discourage some researchers from sharing code, it is clear that the barriers in the case of MIMIC-III are not technical. The unique combination of open code with publicly accessible data allows for the creation of fully executable studies with diligent audit trails, and it would behoove researchers to adopt these approaches.

## Acknowledgments

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.

## Author contributions

AEWJ and TJP collaborated to build the MIMIC code repository. All authors contributed to the paper.

## Competing interests

The authors have no competing interests to declare.

## References

1. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(1):452–4.
2. Gruber WH, Powell AC, Torous JB. Healthcare The power of capturing and using information at the point of care. *Healthcare* [Internet]. 2016;(January):0–1. Available from: <http://dx.doi.org/10.1016/j.hjdsi.2016.09.004>
3. Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. MIMIC-iii, a freely accessible critical care database. *Scientific data*. 2016;3.
4. MIMIC code repository. <https://github.com/MIT-LCP/mimic-code>;
5. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. Best practices for scientific computing. *PLOS Biology* [Internet]. 2014 Jan;12(1):e1001745. Available from: <http://dx.doi.org/10.1371/journal.pbio.1001745> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?art>
6. Knaus WA. APACHE 1978-2001: the development of a quality assurance system based on prognosis: milestones and personal reflections. *Archives of Surgery*. 2002;137(1):37–41.
7. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos CA P G and Sirio, et al. The APACHE III iII prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619–36.
8. Le Gall J, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) based on a European/North American Multicenter Study. *JAMA*. 1993;270(24):2957–63.
9. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas a, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care

- unit. ICU Scoring Group. JAMA [Internet]. 1996 Sep;276(10):802–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17166357>
10. Johnson AE, Kramer AA, Clifford GD. A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Critical Care Medicine*. 2013;41(7):1711–8.
  11. Vincent J-L, Moreno R, Takala J, Willats S, De Mendoca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*. 1996;22:707–10.
  12. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas a, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. JAMA [Internet]. 1996 Sep;276(10):802–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17166357>
  13. Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, et al. Model for end-stage liver disease (mELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91–6.
  14. Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P. Acute renal failure—definition, outcome measures, animal models, fluid therapy and information technology needs: The second international consensus conference of the acute dialysis quality initiative (aDQI) group. *Critical care*. 2004;8(4):R204.
  15. Kellum JA, Mehta RL, Angus DC, Palevsky P, Ronco C. The first international consensus conference on continuous renal replacement therapy. *Kidney international*. 2002;62(5):1855–63.
  16. Mehta RL, Kellum JA, Shah SV, Molitoris BA, Ronco C, Warnock DG, et al. Acute kidney injury network: Report of an initiative to improve outcomes in acute kidney injury. *Critical care*. 2007;11(2):R31.
  17. CKD-MBD Work Group KDIGO (KDIGO), others. KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of chronic kidney disease-mineral and bone disorder (cKD-mBD). *Kidney international Supplement*. 2009;(113):S1.
  18. Ghassemi M, Wu M, Hughes MC, Szolovits P, Doshi-Velez F. Predicting intervention onset in the ICU with switching state space models. *AMIA CRI*. 2017;
  19. Torio CM AR. National inpatient hospital costs: The most expensive conditions by payer, 2011. hCUP statistical brief #160. Agency for Healthcare Research and Quality, Rockville, MD [Internet]. 2013 Aug; Available from: <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf>
  20. Torio CM MB. National inpatient hospital costs: The most expensive conditions by payer, 2013. hCUP statistical brief #204. Agency for Healthcare Research and Quality, Rockville, MD [Internet]. 2016 May; Available from: <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions.pdf>
  21. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*.

2016;315(8):801–10.

22. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*. 2016;315(8):762–74.

23. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the united states: Analysis of incidence, outcome, and associated costs of care. *Critical care medicine*. 2001;29(7):1303–10.

24. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the united states from 1979 through 2000. *New England Journal of Medicine*. 2003;348(16):1546–54.

25. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P, et al. Identifying patients with severe sepsis using administrative claims: Patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Medical care*. 2014;52(6):e39.

26. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical care*. 1998;36(1):8–27.

27. Steiner C, Elixhauser A, Schnaier J. The healthcare cost and utilization project: An overview. *Effective clinical practice: ECP*. 2001;5(3):143–51.

28. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in iCD-9-cM and iCD-10 administrative data. *Medical care*. 2005;1130–9.

29. Walraven C van, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical care*. 2009;626–33.

30. Pérez F, Granger BE. IPython: A system for interactive scientific computing. *Computing in Science and Engineering* [Internet]. 2007 May;9(3):21–9. Available from: <http://ipython.org>

31. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. 2016;87.

32. Hsu DJ, Feng M, Kothari R, Zhou H, Chen KP, Celi LA. The association between in-dwelling arterial catheters and mortality in hemodynamically stable patients with respiratory failure: A propensity score analysis. *CHEST Journal*. 2015;148(6):1470–6.

33. Longo DL, Drazen JM. Data sharing. *Mass Medical Soc*; 2016.

34. Perez-Riverol LAW Yasset AND Gatto. Ten simple rules for taking advantage of git and gitHub. *PLOS Computational Biology* [Internet]. 2016 Jul;12(7):1–11. Available from: <https://doi.org/10.1371/journal.pcbi.1004947>

35. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: The wall must come down. *Jama*. 2013;309(13):1355–6.

36. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *Mass*

Medical Soc; 2017.