A Critical Look at a Critical Care Dataset: MIMIC-IV's Construction, Contents, & Consequences

Pınar Barlas
Faculty of Information & Media Studies
The University of Western Ontario
London, ON, Canada
pbarlas3@uwo.ca

Abstract

MIMIC (Medical Information Mart for Intensive Care) is one of the largest, most commonly-used, freely available datasets containing intensive care unit data. I conduct denotative, connotative, and deconstructive readings of the MIMIC-IV dataset through an analysis of the data sources, dataset structure, and the process for getting access to the data, as well as documents and concepts related to the dataset. As a result, I demonstrate that the MIMIC-IV dataset requires more documentation, including an expansion of the existing descriptions, in order to ensure the data is used appropriately and allow for maximum benefit. I make recommendations for future users of the MIMIC-IV dataset, creators of datasets in general, and researchers in the Critical Data Studies field based on my findings.

Introduction

Healthcare is an important domain where technologies such as artificial intelligence (AI) can make a considerable difference in patients' lives, sometimes in life-and-death situations. AI tools have been promoted as a means to help physicians speed up administrative tasks and streamline aspects of diagnosis and treatment, so that they can spend more time with a patient (or see more patients in the same time frame) (Topol 2019). For example, a transcription system called Whisper, and other tools based on it, are being used by medical centers in patients' consultations with doctors—even though its creator OpenAI warned not to do so in high-risk domains.[1]

Yet despite the risks associated with using such tools—e.g., inaccuracy in patient files generated in error by the transcription system, missing information such as body language of the patient or the thought process of the doctor due to the doctor abstaining from any manual note-taking, and added risks to patient privacy—hospitals and individual clinicians are increasingly using such tools in their work.[2]

Especially in the absence of policies and other guardrails in place (e.g., standards), protecting patients from the potential harms of tools selected and used by their healthcare providers is up to the creators of said tools. However, as the methods used by these tools get more complex, it can get more difficult to identify how the potentially harmful outputs or patterns come into being in a particular tool. Since many of the contemporary computational methods are based on large datasets, such as machine learning (ML), one way to minimize the risk of harm in technological tools is to focus on the datasets used to train and test such systems. This does not necessarily mean identifying and removing the "bias" in the dataset—though it can often be useful to e.g., fix the unbalanced representation of social subgroups. Making sure training/test datasets do not contribute to harmful outcomes also means curating them thoughtfully and creating ample documentation.

In the current work, I analyze Medical Information Mart for Intensive Care (MIMIC), which is often cited as one of the largest, most commonly-used, freely available datasets. It currently contains intensive care unit data from one hospital covering 14 years, with over 350,000 unique individuals and their combined 600,000+ hospitalizations/stays, and appears to be used for a variety of tasks. For example, data from the third iteration (MIMIC-III) was used to train a classifier to identify patients who were ready to be discharged from intensive care (McWilliams et al. 2019). The fourth and current iteration (MIMIC-IV) appears to be popular as well, with its accompanying data descriptor article (Johnson et al. 2023b) being accessed over 108,000 times[3] and cited more than 1,500 times[4] as of May 2025, approximately two and a half years after publication. A data processing pipeline for MIMIC-IV, published independently of the creators of MIMIC-IV, contains optional code to run "a range of clinical prediction tasks" including those relating to readmission, length of stay, mortality, and phenotype prediction

---

[1] "Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said." AP News. https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bbb14

[2] E.g., "Why family doctors across Canada are turning to AI scribes—and what it means for patients." CBC News. https://www.cbc.ca/news/health/ai-scribe-second-opinion-1.7390574

[3] "MIMIC-IV, a freely accessible electronic health record dataset." Nature/Scientific Data. https://www.nature.com/articles/s41597-022-01899-x

[4] "Johnson: Mimic-IV." Google Scholar. https://scholar.google.com/scholar?cluster=17057590153839370425&hl=en&as_sdt=2005&sciodt=0,5

(Gupta et al. 2022). One interesting study that made use of the MIMIC-IV dataset (v0.4) with various deep learning models found that the models relied on demographic features to make mortality predictions (Meng et al. 2022). Not only that, but the models appeared to rely on racial attributes differently across subpopulations, and the outputs showed disparate treatment between different groups of ethnicity, gender, and age (Meng et al. 2022). As deep learning methods rely on patterns in the data, the correlation between demographic features and mortality rates probably exists in the data; it is therefore likely that the results of this research would replicate to other ML models trained using MIMIC-IV.

As I will argue, although MIMIC is suitable only for a narrow set of purposes due to its contents and construction process, the shortage of more appropriate datasets of similar sizes may lead to its use for less-suitable goals. Therefore, it is imperative that the shortcomings of this dataset are acknowledged as much as its strengths. To this end, in the current work, I analyze MIMIC-IV through the lens of Critical Data Studies, hoping to encourage similar investigations into popular datasets. I build on previous research that emphasizes the creation of data as a process that embeds values, goals, and perspectives into even what is perceived as "raw" data. I focus on identifying the potential social justice implications of the creation and use of this dataset through the three modes of "reading datasets" as described by Poirier (Poirier 2021).

In order to ensure the dataset can be used for the maximum potential benefit, I argue that the MIMIC-IV dataset requires more documentation, including an expansion of the existing descriptions, in order to make it clear who and what the data represents. In the next section, I give an overview of the literature discussing the entanglement of social injustice and data, as well as the role of datasets in health and machine learning. Then in the following section, I describe the MIMIC dataset through a denotative reading. I combine the connotative and denotative readings into the Analysis section, and end with a discussion of recommendations for the field and to others who would use MIMIC-IV in the Conclusion.

Background

A "special communication" to an Internal Medicine journal starts with the assertion that ML promises to avoid biases, and that an algorithm could "objectively synthesize and interpret" data (Gianfrancesco et al. 2018). Though the rest of the work outlines potential sources of biases in ML algorithms, this assertion shows the perception of ML systems in medicine (and beyond): as tools that can separate data analysis from the sociocultural reality of medicine. Perhaps more importantly, that the data itself is free of subjectivities, such that a well-crafted algorithm or the right technique could interpret it objectively.

However, it has been shown by decades of scholarship in the Science and Technology Studies field that data and measurement are "theory-laden," meaning data—through the decisions of what to measure, how to measure, how to label the measurements, and so on—encodes particular values, views, and assumptions (e.g., Bowker and Star 2008; Chin-Yee and Upshur 2019; Feinberg 2017; Iliadis and Russo 2016; Paullada et al. 2021; Scheuerman, Hanna, and Denton 2021). It can also reflect social structures of oppression and systematic inequalities, since different social groups would inherently produce distinct data patterns as a result of their experiences (Richardson, Schultz, and Crawford 2019). Such patterns can be picked up by ML models and amplified, therefore deepening the existing inequality e.g. in healthcare (Obermeyer et al. 2019).

The unwanted patterns or harmful assumptions baked into the data need to be made visible and handled carefully, in order to minimize harm. Scholars have therefore called for research looking into the ethical implications of data and algorithms to look closer at the construction of datasets, at the dimensions that are often left out of data documentation (e.g., Feinberg 2017; Iliadis and Russo 2016; Paullada et al. 2021; Sambasivan et al. 2021; Denton et al. 2021) To encourage such critical analysis, and disclosure of particular details, tools to improve dataset documentation have been growing in number. For example, "Datasheets for Datasets" (Gebru et al. 2021) has been well-received in certain data-centered fields, and is the first framework suggested for datasets being submitted to the popular Neural Information Processing Systems (NeurIPS) conference.[5] These frameworks, and even the attempt to fill them out (Bandy and Vincent 2021), help increase transparency of the dataset construction process and increase awareness of the dataset's strengths and weaknesses.

Likely inspired by this scholarship, researchers have been analyzing the datasets that are used as "benchmarks" in the ML field; that is, the datasets that researchers can use to test their models, and compare their performance to others who have completed the same task on the same dataset. The ML field is increasingly focused on the same benchmark datasets, despite many datasets being created. For example, one analysis revealed that benchmark datasets often come out of one of twelve elite and primarily Western institutions, concentrating the power to define success and performance metrics (Koch et al. 2021). The call to improve upon benchmarks comes from within fields, too. For example, two of the authors listed for the MIMIC-IV dataset have co-authored a Perspective piece on the need to *thoughtfully and systematically* create benchmark datasets for healthcare (Panch et al. 2020). In a different review, authors (which again included some listed for MIMIC-IV as well) examined 7,000 clinical AI papers to find that the majority of the databases used came from high income countries, with more than half coming from either US or China (Celi et al. 2022). The first and last authors were also found to be predominantly male, and data experts rather than clinicians (Celi et al. 2022). The authors say that "the biggest concern now" is that the population represented in the dataset may be the only one that will benefit from the technologies built upon that dataset, furthering inequality between high and low income countries.[6]

---

[5] "Call for Datasets and Benchmarks 2024." NeurIPS. https://neurips.cc/Conferences/2024/CallForDatasetsBenchmarks

[6] "We need to be much more diverse." STAT News.

In the same article, reporting their review, Celi (an author also on MIMIC-IV) says that the algorithms in neighboring hospitals that serve different patient populations can perform differently, with this difference potentially being much bigger "[b]etween an institution from São Paulo and an institution in Boston."[7] This would be called a dataset shift: the under-performance of an ML system due to "a mismatch between the data set with which it was developed and the data on which it is deployed," which can be caused by sudden or slow changes, such as those in technology, population, setting, or behavior (Finlayson et al. 2021). Dataset shifts can sometimes be the source of unfairness in outcomes between groups, exacerbating healthcare disparities (Chen et al. 2023). One investigation into dataset shifts used MIMIC-IV to train and test replicates of two commercial algorithms, and found that the models degraded over time, in some part due to the change in codes used in the hospital systems (move from International Classification of Diseases, ICD-9 to ICD-10) and other changes over time.[8] This outcome was in line with the real-world performance of one of the algorithms.[9] Therefore, it is important for those using the dataset to know its full context, and for those building the dataset to record details about the creation process of the dataset.

In a rare move at the time, researchers examined a publicly-available dataset (often used as a benchmark) not by looking at its contents but its *genealogy*, examining "major texts around [the dataset's] creation and impact" (Denton et al. 2021). The critical historical analysis revealed two patterns that apply to similar benchmark datasets: first, the central belief/argument that more data is always better; and second, that the crowdworkers—whose labor built the dataset—are treated as nameless, faceless, interchangeable beings and do not interact with (and certainly do not benefit from) the outcomes of their work (Denton et al. 2021). Similarly, another study examined the language used in papers discussing datasets to uncover the values associated with the dataset and regarded highly in the broader field of ML (Scheuerman, Hanna, and Denton 2021). These methods would both fall under connotative readings of datasets, as described by (Poirier 2021), and overlap with the methodology I employ in the current paper.

Methodology

Poirier describes the three modes of reading as: "a denotative reading (extrapolating the literal meaning of values in a dataset), a connotative reading (tracing the socio-political provenance of data semantics), and a deconstructive reading (seeking what gets Othered through data semantics and structure)" (Poirier 2021).

As suggested by Poirier, I use the texts describing the dataset (authored by the creators of MIMIC) to conduct a **denotative** reading. I describe the parts of the dataset, including its authors, access, and other contextual elements, as these texts indicate. I use four texts: 1) the standalone documentation for all releases, 2) the data descriptor article for MIMIC-IV, 3) the description on the PhysioNet portal on which the dataset is hosted, and 4) the description on the page for MIMIC on the website of the lab with which most of the authors are affiliated.

The standalone documentation for MIMIC is available without access to the datasets themselves,[10] and contains sections for general information that applies to all versions (such as data sources, glossary of terms, and instructions for getting access to the dataset), as well as version-specific information divided into releases (II-IV). The data descriptor article (Johnson et al. 2023b) with supplementary information about the MIMIC-IV dataset (particularly its construction process) was published shortly before v2.2. Some of the details regarding the dataset construction process and data description are also described on the PhysioNet site for each release.[11] General information for MIMIC as a whole can be found on the MIMIC page of the MIT Lab for Computational Physiology,[12] however I could not find any information about MIMIC on the website for the Beth Israel Deaconess Medical Center (BIDMC).[13]

The **connotative** and **deconstructive** readings are combined into one in my analysis, because as Poirier indicates, it is harder to fully separate these types of readings in practice. Therefore, I group my analysis into themes instead. I first discuss the connotative aspects of each theme—interpreting the potential consequences of the decisions made in the dataset's creation, its historical and political context, and so on—and add on the deconstructive reading by questioning the absences and tensions of the data.

MIMIC-IV: Denotative Reading

The MIMIC datasets contain "deidentified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center."[14] There is a repository[15] that accompanies MIMIC, which aims to be

---

https://www.statnews.com/2022/04/06/machine-learning-health-datasets-global/

[7] "We need to be much more diverse." STAT News. https://www.statnews.com/2022/04/06/machine-learning-health-datasets-global/

[8] "AI gone astray: How subtle shifts in patient data send popular algorithms reeling, undermining patient safety." STAT News. https://www.statnews.com/2022/02/28/sepsis-hospital-algorithms-data-shift/

[9] "Epic's AI algorithms, shielded from scrutiny by a corporate firewall, are delivering inaccurate information on seriously ill patients." STAT News. https://www.statnews.com/2021/07/26/epic-hospital-algorithms-sepsis-investigation/

[10] "Homepage." MIMIC Documentation. https://mimic.mit.edu/docs/

[11] "MIMIC-IV v3.1." PhysioNet. https://physionet.org/content/mimiciv/3.1/

[12] "MIMIC." Laboratory for Computational Physiology. https://lcp.mit.edu/mimic

[13] "Beth Israel Deaconess Medical Center." BIDMC of Boston. https://www.bidmc.org/

[14] "About MIMIC." MIMIC Documentation. https://mimic.mit.edu/docs/about/

[15] "MIMIC Code Repository (MIT-LCP/mimic-code)." GitHub. https://github.com/MIT-LCP/mimic-code. https://doi.org/10.5281/zenodo.6818823

"a central hub for sharing, refining, and reusing code" used for analysis of the database (Johnson et al. 2018). The data in MIMIC-IV now spans 2008-2022 as of v3.0, though the initial release had files up to 2019 (Johnson et al. 2023c). The latest version is MIMIC-IV v3.1, first released in August 2019 and last updated 11 October 2024 (Johnson et al. 2023a).

The dataset is hosted on the PhysioNet platform (Goldberger et al. 2000), which was established and continues to be funded by subdivisions of the US government's Department of Health and Human Services.[16] The PhysioNet platform contains: an archive of data for biomedical research (such as MIMIC-IV[17]), a library of software for various types of analyses of physiological data, and a collection of educational materials for exploring and analyzing such data.

Five out of the six authors of the dataset listed on PhysioNet have a Massachusetts Institute of Technology (MIT) affiliation. One of those authors has a secondary affiliation at the Beth Israel Deaconess Medical Center, and the sixth author has affiliations at both BIDMC and Harvard Medical School. At the time of writing, of the five authors with the MIT affiliation, four are listed under the People page of the Lab for Computational Physiology,[18] while the fifth indicates on their personal website that they were a Research Affiliate for the lab at one time.[19]

Data sources

The data is collected from Beth Israel Deaconess Medical Center (BIDMC) in Boston, MA (USA), a teaching hospital associated with Harvard Medical School. A large portion of the data in MIMIC-IV comes from "routine clinical care and related activities such as monitoring, provider orders, and billing" (Johnson et al. 2023c), and is stored in a "a custom hospital wide EHR [electronic health records] and an ICU [intensive care unit] specific clinical information system."[20] The latter is the MetaVision clinical information system, which is a bedside monitor, specifically made for intensive/critical care and used in the BIDMC critical care units (Johnson et al. 2023c). "Complementary datasets"—such as Diagnosis Related Groups, the International Classification of Diseases, and the Healthcare Common Procedure Coding System—as well as "state death records" were integrated into the MIMIC-IV dataset to facilitate data interpretation (Johnson et al. 2023c).

It is noted in the data descriptor article, as well as the PhysioNet page, that the Institutional Review Board at the BIDMC granted a waiver of informed consent and approved the sharing of data as a resource for research. Similar to the statements on the PhysioNet credentialing application page, the data is deemed de-identified and "therefore not subject to the HIPAA Privacy Rule restrictions on sharing PHI."[21]

Dataset structure

The MIMIC-IV database is split into modules emphasizing data provenance. The official documentation states that MIMIC-IV is comprised of five modules,[22] although only two are included in the base MIMIC-IV v3.1 dataset available on PhysioNet: the *hosp* module (data from the hospital-wide EHR) and the *icu* module (data from MetaVision within the intensive care units). The *note* module, which contains "deidentified free-text clinical notes," is "currently not publicly available and the structure is subject to change" according to the official documentation.[23] However, the data descriptor article indicates that the *note* module is available as a separate project[24] on PhysioNet (Johnson et al. 2023c). This paper discusses the aforementioned three modules.

The remaining two modules listed on the official documentation page, but not mentioned in the data descriptor article, are available as separate releases and discussed in the official documentation: the *ed* module (from the emergency department) and the *cxr* module (containing chest x-rays).[25] Although missing from most lists in the documentation, there is a sixth called the *ecg* module[26] containing diagnostic electrocardiograms.

***hosp* Module** This module contains data derived from the hospital wide EHRs about the patient's demographics/other aspects, laboratory measurements, microbiology cultures, medication prescription and administration, hospital billing information, online medical record data, service-related information, and information about the "providers" (clinicians) and their orders. While most of the data appears to have been collected in the hospital, some come from "outside of an inpatient encounter" (Johnson et al. 2023c). The latter are referred to as "baseline" values "from before a patient's hospitalization" and are likely the result of visits at "BIDMC affiliated institutes" (Johnson et al. 2023c).

***icu* Module** This module contains data from the bedside monitors, the MetaVision system. The dataset includes information on: intravenous and fluid inputs and their ingredients, "patient outputs," procedures, other events and information, as well as the "caregiver" (clinicians, "who documented data in the [system]"[27]). The tables include different timestamps, indicating the time at which the measurement

---

[16] "About." PhysioNet. https://physionet.org/about/#funding

[17] "MIMIC-IV v3.1." PhysioNet. https://physionet.org/content/mimiciv/3.1/

[18] "People." Laboratory for Computational Physiology. https://lcp.mit.edu/people

[19] "Homepage." Lucas Bulgarelli. https://lbulgarelli.github.io/

[20] "MIMIC-IV v3.1." PhysioNet. https://physionet.org/content/mimiciv/3.1/

[21] "PhysioNet Credentialing Application." PhysioNet. https://physionet.org/credential-application/ (only available when logged in).

[22] "MIMIC-IV documentation." MIMIC Documentation. https://mimic.mit.edu/docs/iv/

[23] "Note." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/note/

[24] "MIMIC-IV-Note: Deidentified free-text clinical notes v2.2." PhysioNet. https://physionet.org/content/mimic-iv-note/2.2/

[25] "MIMIC-IV documentation." MIMIC Documentation. https://mimic.mit.edu/docs/iv/

[26] "ECG." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/ecg/

[27] "caregiver." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/icu/caregiver/

was “relevant” and when they were “validated by bedside staff” (Johnson et al. 2023c). The Core Concepts page[28] of the official documentation dives deeper into the details of what these timestamps mean. The variables *charttime* and *storetime* refer respectively to when a particular observation was “charted” versus when it was recorded in the database. They state that in the ideal case, the *charttime* indicates the actual time the measurement was taken, and *storetime*, soon after that timestamp, indicates the time that the care staff “validated” and recorded the measurement into the database.

However, in some cases this is not true. For example, *charttime* is said to be reflecting a habit from when nursing staff would record their observations on paper charts, and note the time into hourly blocks. Following this method, “in many cases” an observation made at e.g. 04:23 would be charted as occurring at 04:00 (at the top of the hour, in the 4:00-5:00 block). Lower on the same page, however, the authors say that *charttime* is “usually exactly the time of measurement” with continuous vital signs (e.g. heart rate).

The observations made do not seem to be recorded in the database immediately after the measurement; the authors point out that “All observations in the database must be validated before they are archived into the patient medical record.” The validation is explained as “impl[ying] that the care staff believed the measurement to be an accurate reflection of the patient status at the given *charttime*.” Although it is acknowledged that *storetime* (the time of validation and recording of the data) “logically [] occurs after *charttime*” (the measurement), e.g. half an hour later, there is a short description of cases where *storetime* may come before *charttime*. An example given is that the care staff may make and validate the observation “slightly before” (e.g. at 03:50) the *charttime* recorded (04:00). In this case it is implied that the care staff made the assumption that the observation of the patient, such as vitals, will be the same at some point in the future (10 minutes later, in this example).

***note* Module** This module contains free-text clinical notes of two types. In the *radiology* table, there are semi-structured reports associated with radiography imaging, including x-ray, computed tomography, magnetic resonance imaging, and ultrasound.

In the *discharge* table, there are Discharge summaries, which are “long form narratives”[29] that give an “overview a patient’s history and course throughout a given hospitalization” (Johnson et al. 2023c). While the notes still include sections such as “chief complaint, history of present illness, past medical history, brief hospital course, physical exams, and discharge diagnoses,” the “Social History” and “Discharge Instructions” sections were removed for deidentification purposes (Johnson et al. 2023c).

Dataset Construction Process

MIMIC-IV’s construction process is described as a process in which the data for “all patients who were admitted to either the emergency department or an ICU [at the BIDMC] between 2008-2019[*sic*]” was filtered, excluding patients that were younger than 19 on their first visit and those who “were on a known list of individuals requiring enhanced protection” (Johnson et al. 2023c). The authors note in the data descriptor article that they merged certain medical record numbers as these referred to the same individual, but that they “sought to minimize processing to align the published data as closely to the data recorded during clinical practice” (Johnson et al. 2023c). Similarly, they state in the PhysioNet page description that the preparation of the data was simply “reorganized,” and that “[i]mportantly, data cleaning steps were not performed, to ensure the data reflects a real-world clinical dataset.”[30] However, in the same paragraph, the authors also mention that they removed “audit trails.”

In order to deidentify the data, one of the measures taken was to shift all dates to mask potentially identifying values. The dates are internally consistent—i.e. time-related measurements for a patient can be compared—but randomly distributed in the future so that different patients cannot be compared on the basis of time.[31] The authors note on the PhysioNet page that “all patients over 89 have been grouped together into a single group with value 91.” This is not mentioned in the data descriptor article.

Access to the Dataset

In order to get access to the MIMIC-IV dataset, after creating an account on the PhysioNet platform, interested individuals must “apply” to become a “credentialed user.”[32] The instructions state: “[the account becoming “credentialed”] means that you must formally submit your personal details for review, so that we can confirm your identity.” The credentialing process asks the user to add “an institutional/academic email address” and an ORCID iD, if available, in order to “expedite” the verification process. [33] The process requires that the user share the name and contact information of a reference—for students and postdocs, it must be their supervisor, otherwise the user “may list a colleague.” The form also asks the user to indicate their “Researcher category,” which spans various student roles as well as researcher roles within the industry and academia, including the possibility of registering as an independent researcher. Then, the user is asked to describe their Research Topic, including the specific dataset(s) they intend to use, and how they plan to use the data. There is a note to include the name and number of the course, if the data is going to be used for a class. Lastly, the user must agree to the terms and conditions, which emphasize that the clinical datasets on PhysioNet originally contained “protected health information []

---

[28] “Core concepts.” MIMIC Documentation. https://mimic.mit.edu/docs/iv/about/concepts/

[29] “Discharge.” MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/note/discharge/

[30] “MIMIC-IV v3.1.” PhysioNet. https://physionet.org/content/mimiciv/3.1/

[31] “Core concepts.” MIMIC Documentation. https://mimic.mit.edu/docs/iv/about/concepts/

[32] “PhysioNet Credentialing.” PhysioNet. https://www.physionet.org/settings/credentialing/ (only available when logged in).

[33] “PhysioNet Credentialing Application.” PhysioNet. https://physionet.org/credential-application/ (only available when logged in).

as defined by HIPAA," and that data will only be released to "legitimate researchers" despite being de-identified after "scrupulous attention" on the part of the data providers.

In addition to applying to get their identity verified, the user needs to submit evidence of having completed a "suitable training program in human research subject protections and HIPAA regulations."[34] PhysioNet "recommends" the "Data or Specimens Only Research" course from the Collaborative Institutional Training Initiative (CITI) Program, although the field in Account Settings to submit the proof of training only allows for the CITI program's course. The user is required to upload not the certificate for completion, but the report, which is said to list the modules completed along with dates and scores.

Finally, in order to get access to MIMIC-IV, the user needs to agree to the "PhysioNet Credentialed Health Data Use Agreement 1.5.0."[35] The agreement has 10 items, which include not attempting to identify any individual or institution referenced in the restricted data, not sharing the access with others, and "contribut[ing] the code used to produce [the openly-disseminated] results" to an open repository.

Analysis: Connotative and Deconstructive Readings

In this section, I discuss the cultural and political context of the dataset, performing connotative and denotative readings of the dataset as described in (Poirier 2021).

Generalizability

"Single-source bias," as the name suggest, is the imbalance in performance of a model which develops as a result of the dataset coming from a single system or context. Single-source bias was mentioned as one of the problems in the healthcare technology field in a review from 2022, and cited as the reason models may under-perform on data from other sources (Rajpurkar et al. 2022).

Although it is one of the biggest clinical datasets out there, MIMIC-IV dataset contains data from only one hospital, the BIDMC in Boston, MA, USA. This means that the conclusions that can be drawn from the data are at risk of being unique to the country, state, or city; the segment of the local population that is able to become a patient at the BIDMC; or even to the specific group of people that happened to have been a patient at the BIDMC in the period in question. For example, the patients at the BIDMC may not be representative of Boston—let alone the rest of the country or world—regarding their income level, race/ethnicity, or gender ratio. In fact, the representation and discounted/free treatment of low-income and uninsured patients at BIDMC is lower than other hospitals.[36] Additionally, there is a significant disparity between Black and White patients in readmission after certain medical procedures, and the representation of various non-white racial/ethnic groups (in patients admitted to BIDMC) are between 45%-75% lower than the community.[37] A study analyzing MIMIC-IV reports that the majority of the data comes from White (65.6%) and male (56.1%) patients (Sauer et al. 2022).

In addition, the BIDMC is a teaching hospital, which differs from other public and private hospitals due to its nature as a hospital intended for research and education. Research has found that teaching hospitals have different quality of care and health outcomes as compared to other types of hospitals; for example, teaching hospitals (especially in the U.S.) have been found to have more favorable health outcomes overall, such as lower mortality rates, than non-teaching hospitals (Kupersmith 2005; Burke et al. 2017). Therefore, the findings relating to medical procedures, health outcomes, or other healthcare research may not generalize to patients at non-teaching hospitals.

On their page for MetaVision,[38] iMDsoft feature some research where MetaVision affected healthcare results. In (Levesque et al. 2015), the authors conduct an observational study in a specialized Intensive Care Unit at Paul Brousse University Hospital using MetaVision. The authors find that the introduction of MetaVision shortened the length of stay in the ICU, when controlling for other variables. In (McCambridge et al. 2010), the "academic" or "teaching" community hospital is not named, and neither is its location; however, the study received approval from the ethics review board of a health network with hospitals in various locations around Pennsylvania, U.S. so the study was likely conducted at one of such hospitals. Additionally, the study reports the favorable outcomes that a "health information technology bundle" provided to the patients, of which MetaVision was only one. As implied by the studies, the use of MetaVision may have an impact on the health and experience of patients at the BIDMC intensive care unit, further risking the generalizability of the conclusions that can be drawn from the MIMIC-IV data. Even without MetaVision's (positive) impact on mortality rates or lengths of stay, the presence of a specific information system may be systematically biasing the conditions at the BIDMC intensive care units. The different aspects of the interface, the reliance on various features of the system, and other factors may all be at play in the MIMIC-IV data.

Consent/Refusal

There are no consent mechanisms in any of the information sources about MIMIC-IV. There is a "General Agreement" available on BIDMC's website,[39] which does not indicate that the data could/would be retained and used for research purposes. Considering there are no references to the

---

[34] "CITI Course Instructions." PhysioNet. https://physionet.org/about/citi-course/

[35] "Credentialed Health Data Use Agreement 1.5.0." PhysioNet. https://physionet.org/content/mimiciv/view-dua/3.1/

[36] "Beth Israel Deaconess Medical Center in Boston, MA." US News Best Hospitals Rankings. https://health.usnews.com/best-hospitals/area/ma/beth-israel-deaconess-medical-center-6140013

[37] Ibid.

[38] "MetaVision ICU." iMDsoft. https://www.imdsoft.com/metavision-products/mv-for-intensive-care-adults-neonatal-and-pediatric

[39] "General Agreement." Beth Israel Deaconess Medical Center. https://www.bidmc.org/-/media/files/mr-0446-ip-op-general-agreement.pdf

MIMIC dataset elsewhere on the website, and no mention of the retention/use of data for research purposes in the Frequently Asked Questions and other informational pages for patients, this implies the patients whose data are collected may not be aware of, or have the opportunity to refuse, the data collection. Furthermore, MIMIC-IV has access to data about patients prior to their hospital stay, from institutions affiliated with the BIDMC; it is very likely that the patients did not know this data is available and/or used for research. BIDMC's "Patient's Rights and Responsibilities"[40] does mention the right (#8) to know about and refuse to participate in research. However, the offer for "more information" in this section points the audience to the "human subjects' research" program and protections. Therefore it is possible that this point is meant to only cover research such as medical trials, and not the collection of data for broad research purposes.

Although de-identified, it is important to ensure the patients are informed and are asked for their consent (or the option to refuse) when their data is collected, retained, and used for purposes other than immediate healthcare for which they were admitted to the hospital. As the data comes primarily from the intensive care units and emergency department, the patients in question are even more vulnerable, at times unconscious or otherwise unable to communicate. It is unclear who qualifies as going on the "known list of individuals requiring enhanced protection" that led to MIMIC-IV creators to filter out certain patients from the dataset entirely, at the beginning of the process.

A citizens' jury study in the U.K. found that there was a general concern around lack of transparency in uses of data, and therefore the jury asked that patients were kept informed about uses of their data, as well as being presented with clear ways to opt out (Ford et al. 2020). It is likely this is a sentiment shared by people whose data may be in the MIMIC datasets. Although it is common for the medical field to waive certain rights in the interest of conducting research, it is important to recognize its potential negative impacts. For example, in the piece discussing the Pima Indian Diabetes Dataset (PIDD), Radin discusses the way data collection, use, and reuse reproduce patterns of settler colonialism (Radin 2017). One such way is the storage and use of healthcare data for research, especially when the participants have not explicitly agreed to such uses of their data, and have not been compensated (Radin 2017). Radin also mentions that the legal and political treatment of the Indigenous community meant that the reservation on which they lived functioned as a sort of "natural and perhaps captive laboratory" (Radin 2017, pg. 50). A hospital such as the BIDMC being the closest one, or perhaps the only one to accept a particular insurance, could mean the patients see no other option; therefore, the hospital could in the same way function as a natural laboratory, which the patients in need could not avoid.

In a review authored by 11 scientists—where the first author is affiliated with the Department of Surgery of the BIDMC, the last author is on the MIMIC-IV dataset, and one author is an employee and shareholder at a large pharmaceutical company—the authors argue that the potential of patient reidentification from publicly available medical datasets is a smaller concern than that of slowing ML research "progress" (Seastedt et al. 2022). They specifically point to the MIMIC database and its use as a demonstration of the great potential of data sharing. It should be considered whether this weighing of concerns should be investigated by hearing from people not benefiting from the medical dataset sharing, too.

Qualitative & Unstructured Data

While available with the same requirements as the base MIMIC-IV dataset, the *note* Module is listed as a separate project, requiring additional work to discover, access, and integrate. Moreover, the official documentation still lists the module as "not publicly available," further disincentivizing the use of the data. Researchers have been working on different methods to automatically extract information from clinical notes for at least ten years now (Kim, Riloff, and Hurdle 2015), though the big potential contained in that unstructured data is still going underutilized (Assale et al. 2019).

The information lost (or ignored) when the clinical notes are left out of the analysis can include discussions around whether a particular medicine was kept in the fridge or not, whether the patient brought in their own supply of a particular medicine, and similar facts (Fitzpatrick 2004) which is not—and perhaps cannot be—included in the structured data. Moreover, relying solely on structured data could lead to underestimating services needed to address social determinants of health, and unstructured data such as clinical notes have been found to be more effective in identifying more complex patient cases (Vest et al. 2017). The clinical notes are already stripped of certain information that makes it difficult to consider social determinants of health, namely the exclusion of the Social History section. As this section would contain a lot of potentially identifying information, it cannot be shared; therefore it is all the more important to ensure the rest of the clinical notes are used. An additional issue with reducing patient records to structured data is in potentially removing the patient's lived experiences from the equation. The patient's perspectives on their experiences, including their own descriptions of the symptoms and the embodied aspects associated with the medical conditions, are immensely important for ensuring the "evidence" for the patient's health is interpreted correctly—an existing challenge for the automated analysis of structured clinical data (Chin-Yee and Upshur 2019).

It is possible that the notes indicate errors in the automated measurements, or other important information about the structured data they are meant to be connected to. This is alluded to in the Usage Notes section for the MIMIC-IV page on PhysioNet,[41] where the authors acknowledge

---

[40] "Patients' Rights and Responsibilities." Beth Israel Deaconess Medical Center. BIDMC of Boston. https://www.bidmc.org/patient-and-visitor-information/patient-information/preparing-for-your-stay/patients-rights-and-responsibilities

[41] "MIMIC-IV v3.1." PhysioNet.

that there may be "implausible values" as a result of the "archival process," and that the data reflects "the idiosyncrasies" of routine clinical practice. There is a recommendation to "follow best practice guidelines" but no indication of what the relevant guidelines may be for the issues present in the dataset. Similarly, there is a variable called warning in the *chartevents* and *datetimeevents* tables of the *icu* module, which "specifies if a warning for this observation was manually documented by the care provider"[42]—but there is no further information on what the warning may be about, or where it would be documented. It is possible that this correlates to information in the *note* module for that event. Lastly, researchers have found that clinicians, even with templates that had a strict structure, interpreted and appropriated data input tools to suit their needs—e.g. leaving fields blank or adding information to a different location than expected by the template (Fitzpatrick 2004). The authors note that these observations of paper input forms indicate a need for digital input forms to be integrated by keeping the context of the daily clinical practice in mind (Fitzpatrick 2004). In a similar fashion, the doctors at BIDMC may have been using the clinical notes field as a place to enter data which pertains to another field, but the interface's design did not allow for it to be input.

Habits & Instincts

The "charttime" and "storetime" timestamps available for measurements, as well as for both types of clinical notes in the *note* module, are inconsistent at best. The measurements taken by the clinicians at a particular time may be stored in the database as having been taken earlier or later than reality. Furthermore, there may be cases where the values are "verified" as being accurate in the (near) future, without checking whether this indeed matched reality. Therefore, the values may be potentially inaccurate for the specific time indicated—while this is kept in mind when a clinician interprets the notes of a nurse, as they know the nurses' habits to record the measurement at the top of the hour, such nuances are not easily translated into an ML model training process where the specific timestamps are used. Especially in an intensive care unit, where there may be fast-moving conditions and illnesses, these 'quirks' relating to the timestamps can be problematic when drawing conclusions from the data. Additionally, the "validation" process is not explained further in the MIMIC-IV documents, and so what it involves (repeated measurements, or perhaps a visual inspection with a 'gut feeling') is unclear.

The instinct that the practitioners develop over the course of their career is also important data and should be considered. For example, researchers found that when using observational health data from two large hospitals in Boston, Massachusetts, laboratory tests (whether they were ordered and their timing) were more closely correlated with odds of survival than the results of said tests (Agniel, Kohane, and Weber 2018). Referring to the complexities inherent in observational data, the authors also noted the importance of considering healthcare processes in the analysis (rather than treating them as 'noise'), as well as avoiding the assumption that the data is randomly distributed, since "doctor and patient behaviors contribute to healthcare processes in purposefully biased ways" (Agniel, Kohane, and Weber 2018, pg. 6). Similarly, in a recent study, researchers tested an early warning system (named COmmunicating Narrative Concerns Entered by Resident Nurses) which uses machine learning models trained on nurses' behavior, such as increased frequency monitoring and note-taking, to predict patients at risk of deterioration (Rossetti et al. 2021). This indicates the importance of keeping clinician behavior in datasets, in addition to other data.

Audit trails, which were removed from the original collected data before releasing MIMIC-IV, refer to the metadata that record user actions such as access to a patient's record and any operations done to the record (e.g., Cruz-Correia et al. 2013). The removal of audit trails could have taken away valuable information regarding the human interactions with the information systems/data. In fact, some researchers have argued that "[audit trail] data—alone and in combination with other datasets—offer a new source of insights, which cannot be gleaned from claims data or clinical data, to support health services research and those studying healthcare processes and outcomes" (Adler-Milstein et al. 2020).

Demographics

The *gender* variable in the *patients* table is described as "the genotypical sex of the patient,"[43] with no further descriptions. In the data descriptor article, this is referred to as "the patient's administrative gender" (Johnson et al. 2023c), presumably referring to the value coming directly from the administrative records. However, it is unclear how it was collected or determined in the first place; the variable label indicates the social identity of the patient, while its description indicates a genetic element (but every patient being subjected to genetic testing for their sex is implausible). Though chromosomal and hormonal gender of a person may be deemed their biological sex in research, this is at odds with how sex (and therefore gender, in cisnormative social structures) is determined at birth and recorded into the state's information infrastructures (Kessler 1990), which complicates the "gender" variable in the MIMIC dataset further. There are different implications for the data, depending on whether this information is determined from the patient's state-issued identification documents, self-declaration from filling out a form, the intake professional's assumptions based on the patient's appearance, or otherwise. Sex and/or gender is often considered an influencing factor in medical research, and as such, the accuracy of the data about a patient's sex and/or gender is important. Additionally, how intersex and/or transgender patients are represented within the BIDMC system and/or MIMIC-IV is unclear. Intersex and trans people are already more likely to experience health inequalities (Zee-

---

https://physionet.org/content/mimiciv/3.1/

[42] "chartevents." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/icu/chartevents/

[43] "patients table." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/hosp/patients/

man et al. 2019), and the use of this dataset for gender-related conclusions or technologies stands to worsen such inequalities for intersex and trans people.

There are variables in the *admissions* table labeled *language*, *marital_status*, *ethnicity*, and *insurance* which are grouped together in the documentation's Detailed description section.[44] Altogether, these variables are described as "information about patient demographics," and the only further description given is a note that the values "may change from stay to stay" as they are documented for each admission. Interestingly, the *ethnicity* variable is labeled *race* instead in the summary table at the top of the page; it is unclear which label is used in the actual dataset. The variations in self-reported race can contribute to unfairness between communities too, since mainstream conversations about racial categories can change the public consciousness and infrastructures recording demographics (Chen et al. 2023). Ancestry has been suggested as an accompanying attribute, which could provide additional information with respect to the probability of certain genetic traits and disorders (Chen et al. 2023). The debate on whether or not to consider the race of the patient in healthcare has slowly evolved into more nuanced discussions on *how* to use race over the last 20 years, as racial identity correlates with social realities and genetic variation which may both impact biological health (Vyas, Eisenstein, and Jones 2020).

Age is another factor often discussed in medical research, but the ages of all patients over 89 have been replaced with the same value in MIMIC-IV. Two possibilities for why this was done are: that age loses some of its importance as a factor beyond a certain age; and there may be a higher possibility of re-identification at higher ages. However, the authors of MIMIC-IV do not give any explanations for why they treated age in this way.

Insufficient Documentation

Although the authors claim that they did not perform any "data cleaning steps," it can be argued that merging records, de-identifying the data, and removing audit trails all count as data cleaning. This invites the question of whether it is possible that the authors performed other steps which they did not perceive as data cleaning (or worth mentioning). Still, there are some further steps for data cleaning that people using the dataset would need to conduct on their own, which some scholars recognized as a potential issue. In order to standardize a portion of the process, they released code which they termed a "processing pipeline" for MIMIC-IV (Gupta et al. 2022).

There are also some confusing and conflicting statements throughout the documents, including the PhysioNet page, the official documentation, and the data descriptor article. For example, the official documentation homepage for MIMIC-IV states that "All patients across all datasets are in the *hosp* module," where the *hosp* module is briefly described as "hospital level data for patients." However, it is also said later in the paragraph that "not all ED [module] patients have hospital data," with no clarification of this contradiction. Other issues, such as a lack of possible values for each variable, or different names for the same concept (e.g. "caregivers" label referring to clinicians) further confuse potential users of the dataset. While some of these issues may be worked out through enough familiarity with the dataset, it also invites assumptions and errors that could easily be cleared up by additional documentation from the dataset creators.

Access

There are various references to the use of PhysioNet resources (e.g. MIMIC-IV and other datasets) for use in education. Considering education is where a lot of future developers will build their habits relating to dataset use, it is important to consider what values and assumptions may be built into the datasets, even through the process to gain access.

The training that the platform requires is described in very general terms, however the system itself only accepts the CITI training program. This may discourage (or even disallow) researchers outside the US from gaining access to datasets in PhysioNet. The PhysioNet platform also has various requirements in place to ensure the individuals that gain access to the datasets they host are "legitimate researchers." This legitimacy seems to be satisfied by having a reference to vouch for oneself and an institutional email. There appear to be no limitations on, and no guidelines for, the use of the datasets in research. Although some uses are implicitly and explicitly discouraged due to dataset structure (e.g. cross-patient analyses in MIMIC-IV), it appears to be the responsibility of the researcher to gauge the suitability of the dataset for the analysis based on the documentation provided. There is a small space when becoming a credentialed user on PhysioNet to discuss the dataset(s) the user intends to use, but this is a one-time requirement and (in theory) not passed on to the creators of the specific dataset. There is implicit trust that the user would honestly disclose the intended use of the datasets to which they were planning on getting access, although this is in conflict with the distrust implied by the request for a report (with the course name and score) in lieu of a completion certificate for their training on regulations/protections, and the requirement of a reference.

Funding and infrastructures

Lastly, it is important to further highlight the infrastructures that MIMIC-IV is built on (beyond the previous releases of MIMIC), and the funding that makes MIMIC-IV a reality. For example, the PhysioNet platform is managed by members of the Massachusetts Institute of Technology (MIT) Lab for Computational Physiology, with close ties to the Margret and H.A. Rey Institute for Nonlinear Dynamics at BIDMC. Similarly, the authors of the dataset are all associated with either the Lab at MIT or the BIDMC. Therefore, it is likely that the authors of the MIMIC dataset manage the platform that it is hosted on. PhysioNet was established, and continues to be funded, by the US government. The MetaVision

---

[44] "admissions." MIMIC Documentation. https://mimic.mit.edu/docs/iv/modules/hosp/admissions/

software is made by iMDsoft, headquartered in Israel.[45] It is a subsidiary of N. Harris Computer Corporation, which is in turn owned by Constellation Software Inc., a public software company in Canada. While the data may come only from Boston, MA, USA, the funds and infrastructure of the dataset spans multiple countries and continents. The involvement of other countries in the provenance of the data must be acknowledged and its impacts considered, especially given the role technology is playing in Israel's genocide of Palestine.[46]

Conclusion

The fourth iteration of the MIMIC dataset, MIMIC-IV continues to be a valuable resource for machine learning researchers and developers working with healthcare applications of data-driven methods. Below, based on my analysis of the MIMIC dataset, I give brief recommendations first to those making use of MIMIC-IV in the future, then to dataset creators, and lastly to researchers—in critical data studies and beyond.

For future MIMIC-IV dataset users: The dataset contains data which may be significantly different from other hospitals, even those from the same city. Therefore, when drawing conclusions or building models using this dataset, it would be fitting to test the conclusion or model with other data, especially those from the context in which the tool or intervention may be deployed. Take a closer look at the *note* module and consider working with it in tandem with the structured data. Ensure the models or conclusions do not rely on very specific timestamps, as the times may not always be accurate.

For (medical) dataset creators, including those updating MIMIC-IV: Consider informing every individual involved in the data creation of the dataset, and allow them to opt out, even if this means the dataset is much smaller. Being explicit and detailed about the context in which the dataset was created would allow the future users of the dataset to interpret the data much more effectively. Even if some data may be harder to work with, due to its unstructured nature, or may appear worthless (such as audit trails), it may still be very valuable to future users. Include information from the host institution/organization, if applicable, relating to the data they are providing. For example, if the demographic data is imported from a different part of the institution, it would be important to reach out to them to find out how it was collected. Document every step of the (pre-)processing of the data, even if it may seem boring or not important.

For researchers in critical data studies and related fields: More investigation is needed into the implications of having authors affiliated with the hospital from which the data originates (even if it is commonplace for medical datasets). Similarly, though a "waiver of consent" is common in observational medical research (as in the case of analyzing MIMIC-IV data), and not required by law or other policy frameworks in place, further investigation into the effect of this on the data subjects should be considered.

Limitations and future work. Further work could look into how the MIMIC-IV dataset is being used, including the types of data analyzed and the purposes of the analyses. Though there is explicit discouragement of the use of the dataset for cross-patient and temporal analyses, it is unclear if this advice is followed by the researchers. Due to the limited space in the current paper, I have limited my analysis to what I felt was most important, but there are many similar paths to extending the connotative and deconstructive readings of the MIMIC dataset. There is also a lot of potential in applying the same methodology (from (Poirier 2021)) to another dataset with similar impact in the field, particularly those that are used most often in different disciplines.

Acknowledgements

I would like to thank Luke Stark for his invaluable support at every stage of this work. I would also like to thank Ben Chin-Yee, Daniel Arauz Nuñez, Takuya Maeda, Jacquie Burkell, Hale Doguoglu, Maxwell Smith, and the anonymous reviewers for their thoughtful feedback.

References

Adler-Milstein, J.; Adelman, J. S.; Tai-Seale, M.; Patel, V. L.; and Dymek, C. 2020. EHR audit logs: A new goldmine for health services research? *Journal of Biomedical Informatics*, 101: 103343.

Agniel, D.; Kohane, I. S.; and Weber, G. M. 2018. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*, k1479.

Assale, M.; Dui, L. G.; Cina, A.; Seveso, A.; and Cabitza, F. 2019. The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records. *Frontiers in Medicine*, 6: 66.

Bandy, J.; and Vincent, N. 2021. Addressing "Documentation Debt" in Machine Learning: A Retrospective Datasheet for BookCorpus. In Vanschoren, J.; and Yeung, S., eds., *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1. Curran.

Bowker, G. C.; and Star, S. L. 2008. *Sorting things out: classification and its consequences*. Inside technology. Cambridge, Mass.: MIT Press, 1. paperback ed., 8. print edition. ISBN 978-0-262-52295-3.

Burke, L. G.; Frakt, A. B.; Khullar, D.; Orav, E. J.; and Jha, A. K. 2017. Association Between Teaching Status and Mortality in US Hospitals. *JAMA*, 317(20): 2105–2113.

Celi, L. A.; Cellini, J.; Charpignon, M.-L.; Dee, E. C.; Dernoncourt, F.; Eber, R.; Mitchell, W. G.; Moukheiber, L.; Schirmer, J.; Situ, J.; Paguio, J.; Park, J.; Wawira, J. G.; Yao, S.; and Data, f. M. C. 2022. Sources of bias in artificial

---

[45] "About." iMDsoft. https://www.imd-soft.com/about-us

[46] Albanese, Francesca. 16 June 2025. "A/HRC/59/23: From economy of occupation to economy of genocide - Report of the Special Rapporteur on the situation of human rights in the Palestinian territories occupied since 1967." UN Human Rights Council. https://www.un.org/unispal/document/a-hrc-59-23-from-economy-of-occupation-to-economy-of-genocide-report-special-rapporteur-francesca-albanese-palestine-2025/

intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health*, 1(3): e0000022.

Chen, R. J.; Wang, J. J.; Williamson, D. F. K.; Chen, T. Y.; Lipkova, J.; Lu, M. Y.; Sahai, S.; and Mahmood, F. 2023. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, 7(6): 719–742.

Chin-Yee, B.; and Upshur, R. 2019. Three Problems with Big Data and Artificial Intelligence in Medicine. *Perspectives in Biology and Medicine*, 62(2): 237–256.

Cruz-Correia, R.; Boldt, I.; Lapão, L.; Santos-Pereira, C.; Rodrigues, P. P.; Ferreira, A. M.; and Freitas, A. 2013. Analysis of the quality of hospital information systems audit trails. *BMC Medical Informatics and Decision Making*, 13(1): 84.

Denton, E.; Hanna, A.; Amironesei, R.; Smart, A.; and Nicole, H. 2021. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society*, 8(2): 205395172110359.

Feinberg, M. 2017. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2952–2963. Denver Colorado USA: ACM. ISBN 978-1-4503-4655-9.

Finlayson, S. G.; Subbaswamy, A.; Singh, K.; Bowers, J.; Kupke, A.; Zittrain, J.; Kohane, I. S.; and Saria, S. 2021. The Clinician and Dataset Shift in Artificial Intelligence. *New England Journal of Medicine*, 385(3): 283–286.

Fitzpatrick, G. 2004. Integrated care and the working record. *Health Informatics Journal*, 10(4): 291–302.

Ford, E.; Oswald, M.; Hassan, L.; Bozentko, K.; Nenadic, G.; and Cassell, J. 2020. Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK. *Journal of Medical Ethics*, 46(6): 367–377.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; III, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gianfrancesco, M. A.; Tamang, S.; Yazdany, J.; and Schmajuk, G. 2018. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*, 178(11): 1544.

Goldberger, A. L.; Amaral, L. A.; Glass, L.; Hausdorff, J. M.; Ivanov, P. C.; Mark, R. G.; Mietus, J. E.; Moody, G. B.; Peng, C. K.; and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23): E215–220.

Gupta, M.; Gallamoza, B.; Cutrona, N.; Dhakal, P.; Poulain, R.; and Beheshti, R. 2022. An Extensive Data Processing Pipeline for MIMIC-IV.

Iliadis, A.; and Russo, F. 2016. Critical data studies: An introduction. *Big Data & Society*, 3(2): 205395171667423.

Johnson, A.; Bulgarelli, L.; Pollard, T.; Horng, S.; Celi, L. A.; and Mark, R. 2023a. MIMIC-IV.

Johnson, A. E. W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T. J.; Hao, S.; Moody, B.; Gow, B.; Lehman, L.-w. H.; Celi, L. A.; and Mark, R. G. 2023b. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1): 1.

Johnson, A. E. W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T. J.; Hao, S.; Moody, B.; Gow, B.; Lehman, L.-w. H.; Celi, L. A.; and Mark, R. G. 2023c. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1): 1. Publisher: Nature Publishing Group.

Johnson, A. E. W.; Stone, D. J.; Celi, L. A.; and Pollard, T. J. 2018. The MIMIC Code Repository: enabling reproducibility in critical care research. *Journal of the American Medical Informatics Association*, 25(1): 32–39.

Kessler, S. J. 1990. The Medical Construction of Gender: Case Management of Intersexed Infants. *Signs: Journal of Women in Culture and Society*, 16(1): 3–26.

Kim, Y.; Riloff, E.; and Hurdle, J. F. 2015. A Study of Concept Extraction Across Different Types of Clinical Notes. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2015: 737–746.

Koch, B.; Denton, E.; Hanna, A.; and Foster, J. G. 2021. Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1.

Kupersmith, J. 2005. Quality of Care in Teaching Hospitals: A Literature Review. *Academic Medicine*, 80(5): 458–466.

Levesque, E.; Hoti, E.; Azoulay, D.; Ichai, P.; Samuel, D.; and Saliba, F. 2015. The implementation of an Intensive Care Information System allows shortening the ICU length of stay. *Journal of Clinical Monitoring and Computing*, 29(2): 263–269.

McCambridge, M.; Jones, K.; Paxton, H.; Baker, K.; Sussman, E. J.; and Etchason, J. 2010. Association of Health Information Technology and Teleintensivist Coverage With Decreased Mortality and Ventilator Use in Critically Ill Patients. *Archives of Internal Medicine*, 170(7).

McWilliams, C. J.; Lawson, D. J.; Santos-Rodriguez, R.; Gilchrist, I. D.; Champneys, A.; Gould, T. H.; Thomas, M. J.; and Bourdeaux, C. P. 2019. Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ Open*, 9(3): e025925.

Meng, C.; Trinh, L.; Xu, N.; Enouen, J.; and Liu, Y. 2022. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Scientific Reports*, 12(1): 7166.

Obermeyer, Z.; Powers, B.; Vogeli, C.; and Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447–453.

Panch, T.; Pollard, T. J.; Mattie, H.; Lindemer, E.; Keane, P. A.; and Celi, L. A. 2020. "Yes, but will it work for my patients?" Driving clinically relevant research with benchmark datasets. *npj Digital Medicine*, 3(1): 87.

Paullada, A.; Raji, I. D.; Bender, E. M.; Denton, E.; and Hanna, A. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11): 100336.

Poirier, L. 2021. Reading datasets: Strategies for interpreting the politics of data signification. *Big Data & Society*, 8(2): 205395172110293.

Radin, J. 2017. “Digital Natives”: How Medical and Indigenous Histories Matter for Big Data. *Osiris*, 32(1): 43–64.

Rajpurkar, P.; Chen, E.; Banerjee, O.; and Topol, E. J. 2022. AI in health and medicine. *Nature Medicine*, 28(1): 31–38.

Richardson, R.; Schultz, J. M.; and Crawford, K. 2019. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online*, 94: 15–55.

Rossetti, S. C.; Dykes, P. C.; Knaplund, C.; Kang, M.-J.; Schnock, K.; Jr, J. P. G.; Fu, L.-H.; Chang, F.; Thai, T.; Fred, M.; Korach, T. Z.; Zhou, L.; Klann, J. G.; Albers, D.; Schwartz, J.; Lowenthal, G.; Jia, H.; Liu, F.; and Cato, K. 2021. The Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) Clinical Decision Support Early Warning System: Protocol for a Cluster Randomized Pragmatic Clinical Trial. *JMIR Research Protocols*, 10(12): e30238.

Sambasivan, N.; Kapania, S.; Highfill, H.; Akrong, D.; Paritosh, P.; and Aroyo, L. M. 2021. “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. Yokohama Japan: ACM. ISBN 978-1-4503-8096-6.

Sauer, C. M.; Dam, T. A.; Celi, L. A.; Faltys, M.; de la Hoz, M. A. A.; Adhikari, L.; Ziesemer, K. A.; Girbes, A.; Thoral, P. J.; and Elbers, P. 2022. Systematic Review and Comparison of Publicly Available ICU Data Sets—A Decision Guide for Clinicians and Data Scientists. *Critical Care Medicine*, 50(6): e581.

Scheuerman, M. K.; Hanna, A.; and Denton, E. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–37.

Seastedt, K. P.; Schwab, P.; O’Brien, Z.; Wakida, E.; Herrera, K.; Marcelo, P. G. F.; Agha-Mir-Salim, L.; Frigola, X. B.; Ndulue, E. B.; Marcelo, A.; and Celi, L. A. 2022. Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digital Health*, 1(10): e0000102.

Topol, E. J. 2019. *Deep medicine: how artificial intelligence can make healthcare human again*. New York: Basic Books, first edition edition. ISBN 978-1-5416-4464-9.

Vest, J. R.; Grannis, S. J.; Haut, D. P.; Halverson, P. K.; and Menachemi, N. 2017. Using structured and unstructured data to identify patients’ need for services that address the social determinants of health. *International Journal of Medical Informatics*, 107: 101–106.

Vyas, D. A.; Eisenstein, L. G.; and Jones, D. S. 2020. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine*, 383(9): 874–882.

Zeeman, L.; Sherriff, N.; Browne, K.; McGlynn, N.; Mirandola, M.; Gios, L.; Davis, R.; Sanchez-Lambert, J.; Aujean, S.; Pinto, N.; Farinella, F.; Donisi, V.; Niedźwiedzka-Stadnik, M.; Rosińska, M.; Pierson, A.; Amaddeo, F.; Network, H.; Taibjee, R.; Toskin, I.; Jonas, K.; Van Der Veur, D.; Allen, O.; Troussier, T.; and De Sutter, P. 2019. A review of lesbian, gay, bisexual, trans and intersex (LGBTI) health and healthcare inequalities. *European Journal of Public Health*, 29(5): 974–980.