Further Anonymizing The Profiles of Patients Presenting to the Emergency Room Using "De-Identified" Data

Members

Sanjiv M. Narayan, Edward Tatchim, Rohan Krishnamurthi

Goal: To identify different clinical profiles ("Phenotypes") of patients who presented to the Emergency room of Harvard Teaching Hospitals, and use k-anonymity, L-diversity and t-closeness to show that structured various generalization schemes can effectively "hide" these profiles and reduce attribute or membership disclosure. For this, we will use real-world medical data, in the form of MIMIC-IV released through MIT and Beth Israel Hospital and covering admissions to their intensive care unit from 2008 to 2019. We can explore hiding of DEI attributes/memberships or other features if we have time ☺ This could be published if we can get it done by 12/02/2024 (I'll wrap up and submit on our behalf).

Data MIMIC-IV is a publicly available database sourced from the electronic health record of the Beth Israel Deaconess Medical Center. Information available includes patient measurements, orders, diagnoses, procedures, treatments, and deidentified free-text clinical notes.[1] This is an extensive data repository as summarized in table 1 below.[1] For this proof of concept study, we used a subset represented in this dataset. https://physionet.org/content/mimic-iv-ecg-ext-icd-labels/1.0.1/ [2]

	Hospital admissions	ICU admissions
Number of stays	431,231	73,181
Unique patients	180,733	50,920
Age, mean (SD)	58.8 (19.2)	64.7 (16.9)
Female Administrative Gender, n (%)	224,990 (52.2)	32,363 (44.2)
Insurance, n (%)		
Medicaid	41,330 (9.6)	5,528 (7.6)
Medicare	160,560 (37.2)	33,091 (45.2)
Other	229,341 (53.2)	34,562 (47.2)
Hospital length of stay, mean (SD)	4.5 (6.6)	11.0 (13.3)
In-hospital mortality, n (%)	8,974 (2.1)	8,519 (11.6)
One year mortality, n (%)	106,218 (24.6)	28,274 (38.6)

Table 1. Demographics for patients admitted to an intensive care unit (ICU) in MIMIC-IV v2.2.

All data are deidentified in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision. Look-up tables were used to randomly assign patients with a unique identifier (subject_id) and hospitalizations with a unique identifier (hadm_id). Dates were perturbed by shifting them using a patient-level offset. The shift ensures that the interval between two time points for a patient is preserved. The authors combined 2 published algorithms to remove PHI from the database and replace it with 3 consecutive underscores ("___").

MIMIC-IV table	Description of change
emar, emar_detail, ingredientevents, omr, poe, poe_detail, pharmacy	New tables only present in MIMIC-IV.
labevents	Added the following columns: `storetime`, `specimen_id`, `ref_range_lower`, `ref_range_upper`, `priority`, and `comments`.
microbiologyevents	Added the following columns: `micro_specimen_id`, `test_seq`, `storedate`, `storetime`, `test_name`, `test_name`, `quantity`, `comments`.
labevents	Added the following columns: `storetime`, `specimen_id`, `ref_range_lower`, `ref_range_upper`, priority, and `comments`.
hcpcsevents, d_hcpcs	Replaced the *cptevents* and *d_cpt* tables.
prescriptions	Columns `starttime` and `endtime` replaced `startdate` and `enddate` as the associated time is now available.
	Columns `drug_name_generic` and `drug_name_poe` were removed.
	Columns `pharmacy_id`, form_rx, and `pharmacy_id` were added.
inputevents	Renamed; equivalent to *inputevents_mv* in MIMIC-III. *inputevents_cv* has been removed.
procedureevents	Renamed; equivalent to *procedureevents_mv* in MIMIC-III.
icustays	The unit-level identifier *icustay_id* has been replaced with the general location-based identifier *stay_id*. *stay_id* is used to identify a period of stay within a single location.

Table 2. Major changes between MIMIC-III v1.4 and MIMIC-IV v2.2.

Proposed Work

In a world where data is the new currency, protecting sensitive medical information has never been more critical. The sheer volume of healthcare data being generated holds incredible potential for breakthroughs in treatment, research, and AI-driven innovation.

Medical data presents unique challenges for privacy engineering, with both unique risks to privacy but also critical utility. While data in many domains can be altered or even reset if needed, medical identifiers such as health status, biometrics and even genetic data cannot (Price and Cohen, Nat Comm 2019). Moreover, while techniques such as k-anonymity and differential privacy operate by generating statistical results from which an individual's attributes are difficult to ascertain, unique attributes are often required for medical therapy. Such challenges are even more acute given the convergence of data to which access is protected under the Health Information Portability and Accountability Act (HIPAA; Mandl and Perakslis, NEJM 2022), and commercial technologies which are often not covered entities under such regulations.

Our goal is to anonymize publicly available medical data, focusing on identifying and protecting quasi-identifiers and attributes that are accessible online. Safeguarding patient privacy while maintaining the utility of healthcare data is critical in providing valuable medical insights without compromising patient identity. To achieve this, we will adhere to various medical regulations and guidelines, ensuring that the anonymization process is both rigorous and compliant.

For implementation, we aim to address the risk disclosure triad: attribute disclosure, identity disclosure, and membership disclosure. These are key concerns when making medical data publicly available and safe for use in predictive AI models.

Survey of Related Work

The anonymization of public medical data is a well-established practice, driven by the need to balance privacy with the data's inherent value for research and education. De-identifying healthcare data is essential for protecting patient privacy and complying with regulations such as HIPAA. Medical data comes in many forms—ranging from quantitative metrics to multimedia representations of medical conditions, diseases, and anatomical features. The diversity of data formats adds complexity to the anonymization process, requiring tailored approaches.

For instance, images may need to be blurred to obscure identifiable features, and voices distorted to mask individuals' identities. The challenge lies in anonymizing the data while preserving clinically relevant information, ensuring that the insights drawn remain valuable without compromising patient confidentiality.

Preliminary Results or Investigation

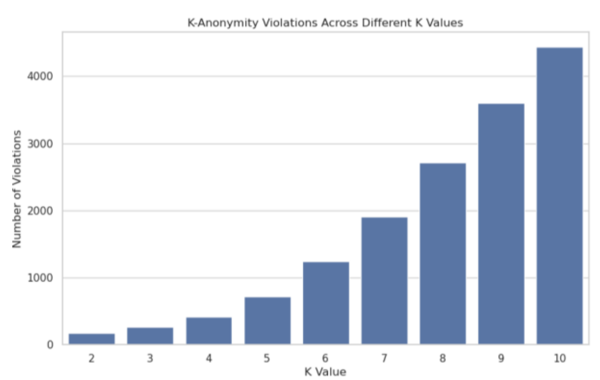
This synthetic healthcare dataset mimics real-world patient records, designed for data science and machine learning practice. Generated using Python’s Faker library, it enables hands-on experience in healthcare analytics without actual patient privacy concerns. This fully synthetic dataset is a secure, privacy-compliant resource for healthcare data analysis, fostering innovation and learning in data science.

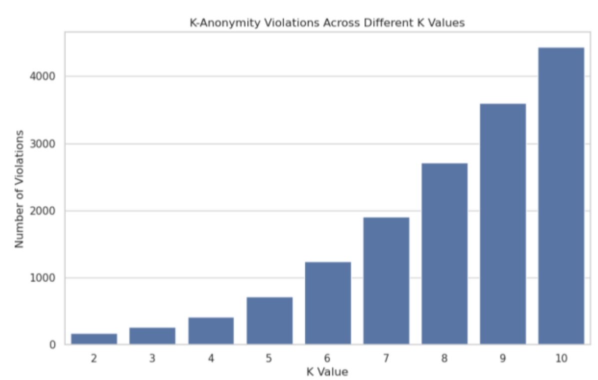
Here's a brief explanation of each column in the dataset -

- **Name:** This column represents the name of the patient associated with the healthcare record.
- **Age:** The age of the patient at the time of admission, expressed in years.
- **Gender:** Indicates the gender of the patient, either "Male" or "Female."
- **Blood Type:** The patient's blood type, which can be one of the common blood types (e.g., "A+", "O-", etc.).
- **Medical Condition:** This column specifies the primary medical condition or diagnosis associated with the patient, such as "Diabetes," "Hypertension," "Asthma," and more.
- **Date of Admission:** The date on which the patient was admitted to the healthcare facility.
- **Doctor:** The name of the doctor responsible for the patient's care during their admission.
- **Hospital:** Identifies the healthcare facility or hospital where the patient was admitted.

- **Insurance Provider:** This column indicates the patient's insurance provider, which can be one of several options, including "Aetna," "Blue Cross," "Cigna," "UnitedHealthcare," and "Medicare."
- **Billing Amount:** The amount of money billed for the patient's healthcare services during their admission. This is expressed as a floating-point number.
- **Room Number:** The room number where the patient was accommodated during their admission.
- **Admission Type:** Specifies the type of admission, which can be "Emergency," "Elective," or "Urgent," reflecting the circumstances of the admission.
- **Discharge Date:** The date on which the patient was discharged from the healthcare facility, based on the admission date and a random number of days within a realistic range.
- **Medication:** Identifies a medication prescribed or administered to the patient during their admission. Examples include "Aspirin," "Ibuprofen," "Penicillin," "Paracetamol," and "Lipitor."
- **Test Results:** Describes the results of a medical test conducted during the patient's admission. Possible values include "Normal," "Abnormal," or "Inconclusive," indicating the outcome of the test.

Quasi-identifiers are attributes like **Age**, **Gender**, **Blood Type** and **Medical Condition** that, when combined, can reveal an individual's identity. Though not unique on their own, their combination increases the risk of re-identification.





Click here to see more related graphs

Contributions of Each Members

Member	**Contributions**
Edwin Figueroa	Exploratory Data Analysis and Data Visualizations
Edward Tatchim	Analyzing related datasets and designing a privacy engineering solution to address risk disclosure triad

Rohan Krishnamurthi	Content Composition & Solution Proposal
Sanjiv Narayan	Project Vision, Scoping and Purpose

Proposed Milestones

1. Analyzing related work
2. Identifying problem statement
3. Defining the importance of a privacy solution
4. Defining the scope of the project
    a. What aspect of privacy engineering do we want to focus on?
5. Background and Context
    a. Identifying and defining medical dataset of interest
    b. Describing existing regulations protecting medical datasets (HIPAA, GDPR, etc)
    c. Describing existing regulations around AI modeling
6. Deep dive into the challenges of ensuring confidentiality in medical data
    a. Vulnerabilities
    b. Threat models
    c. Areas of misuse
7. Identifying privacy engineering techniques
    a. Identifying techniques
    b. Testing techniques
    c. Implementing techniques
8. Identifying regulatory Compliance and Ethical Considerations
    a. Defining regulatory frameworks to support proposed solution
    b. Defining ethical implications of proposed solution
    c. Establishing accountability and auditing systems for proposed solution

Sources
Relevant Work:
https://enlitic.com/blogs/deidentifying-and-anonymizing-healthcare-data/#:~:text=Anonymizing%20healthcare%20data%2C%20on%20the,with%20regulations%20such%20as%20HIPAA.

Dataset curation
Kaggle dataset:
https://www.kaggle.com/datasets/prasad22/healthcare-dataset

Actual Electronic Medical Record Data

https://physionet.org/content/mimic-iv-ecg-ext-icd-labels/1.0.1/

MIMIC-IV abbreviated open source data [2]

Original data from **PhysioNet** [3]

Abstract

MIMIC-IV-ED is a publicly accessible database of over 400,000 emergency department (ED) admissions to the Beth Israel Deaconess Medical Center between 2011 and 2019. Access to MIMIC-IV-ED requires registration on PhysioNet, identity verification, completion of human participant training, and signing of a data use agreement. Here, we have provided an openly available demo of MIMIC-IV-ED containing a subset of 100 patients. Being a derivative dataset, the MIMIC-IV-ED demo is deidentified according to the same standard as MIMIC-IV-ED. The demo is intended to support workshops, educational material, exploration of MIMIC-IV-ED, and other aims enabled by an openly available dataset.

References

1 Johnson, A. E. W. *et al.* MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* **10**, 1 (2023). https://doi.org:10.1038/s41597-022-01899-x

2 Johnson, A. *et al. MIMIC-IV-ED Demo* *https://physionet.org/content/mimic-iv-ed-demo/2.2/*, 2023).

3 Goldberger, A. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101**, e215-e220 (2000).

Title: The ECG As a Personal Digital Fingerprint

Background: There is increasing concern on the ability to re-identify clinical notes from even de-identified data that meets HIPAA approaches. ventricular ejection fraction (LVEF) is an important marker of cardiac health and well-being, but it is not always convenient to obtain serial echocardiograms in patients at risk. We developed a multi-layered deep neural network for the routine 12-lead ECG to predict temporal trajectories of LVEF over time.

Objective: To test the hypothesis that deep learning applied to a single-lead ECG, trained on hundreds of thousands of beats in 12-lead ECGs referenced to LVEF from the echocardiogram, can be used to estimate LVEF trends in a given patient over time.

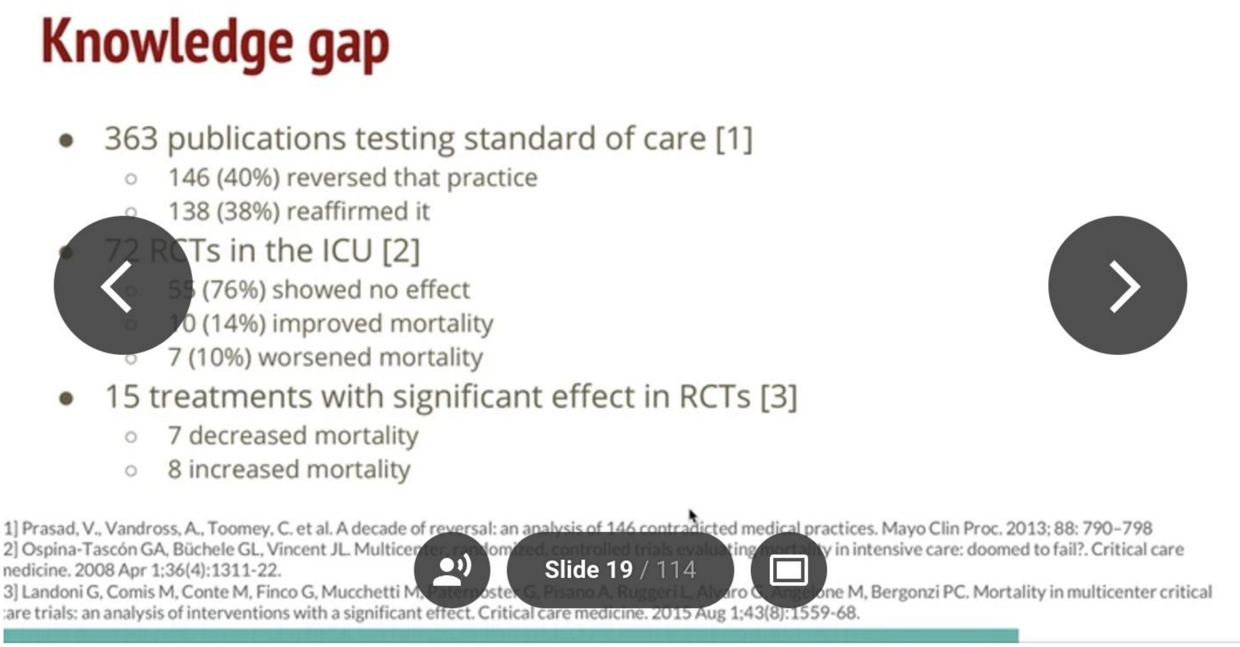
Methods: We extracted *** echocardiogram reports from *** patients who visited the Stanford hospital between May 2014 and August 2018, matched to 12-lead ECG recordings within $\pm$ 14 days. We first developed a DL based 12-lead ECG classifier to classify reduced echocardiographic LVEF. Using this DL-based ECG-LVEF classifier for pre-training, we developed a multi-layered deep convolutional U-Net to reconstruct the 12-lead ECG from a single V3 – V2 bipole. In a hold-out-test set (N=**), we applied the trained U-Net encoder + classifier to predict a) drop in LVEF, and b) absolute LVEF in serial measurements from the same patient.

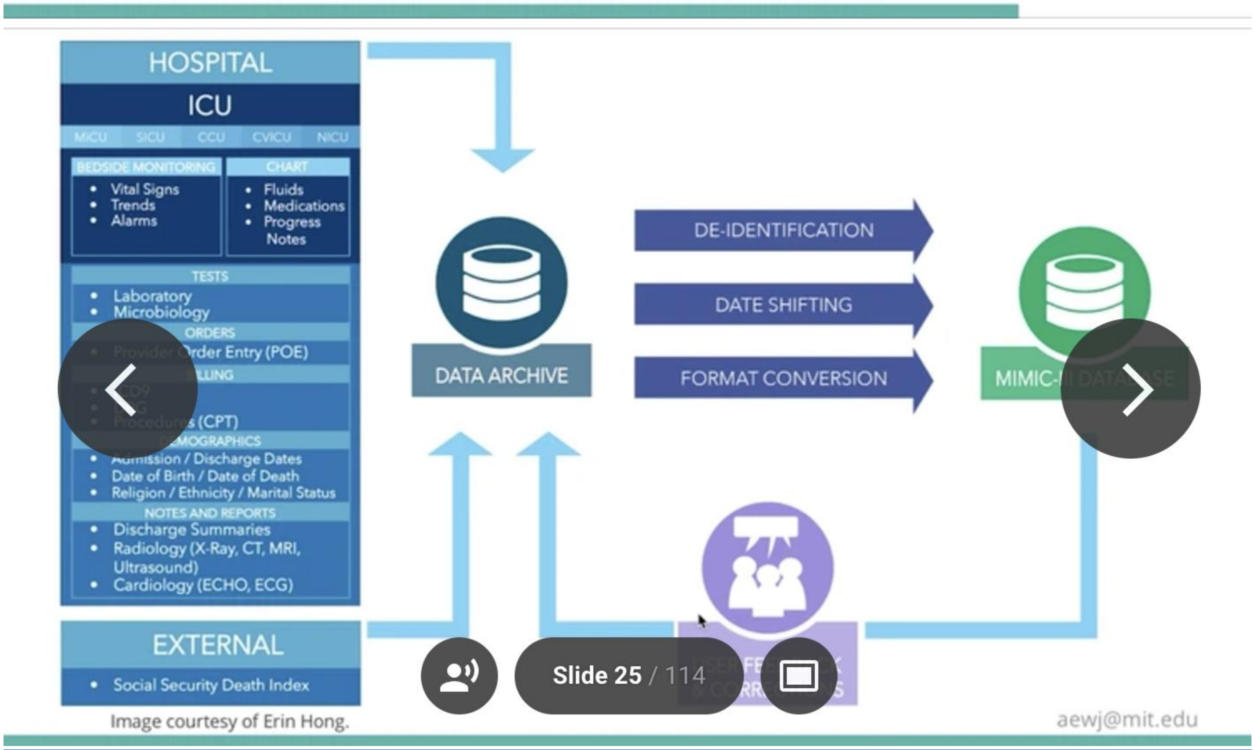
FIGURE **

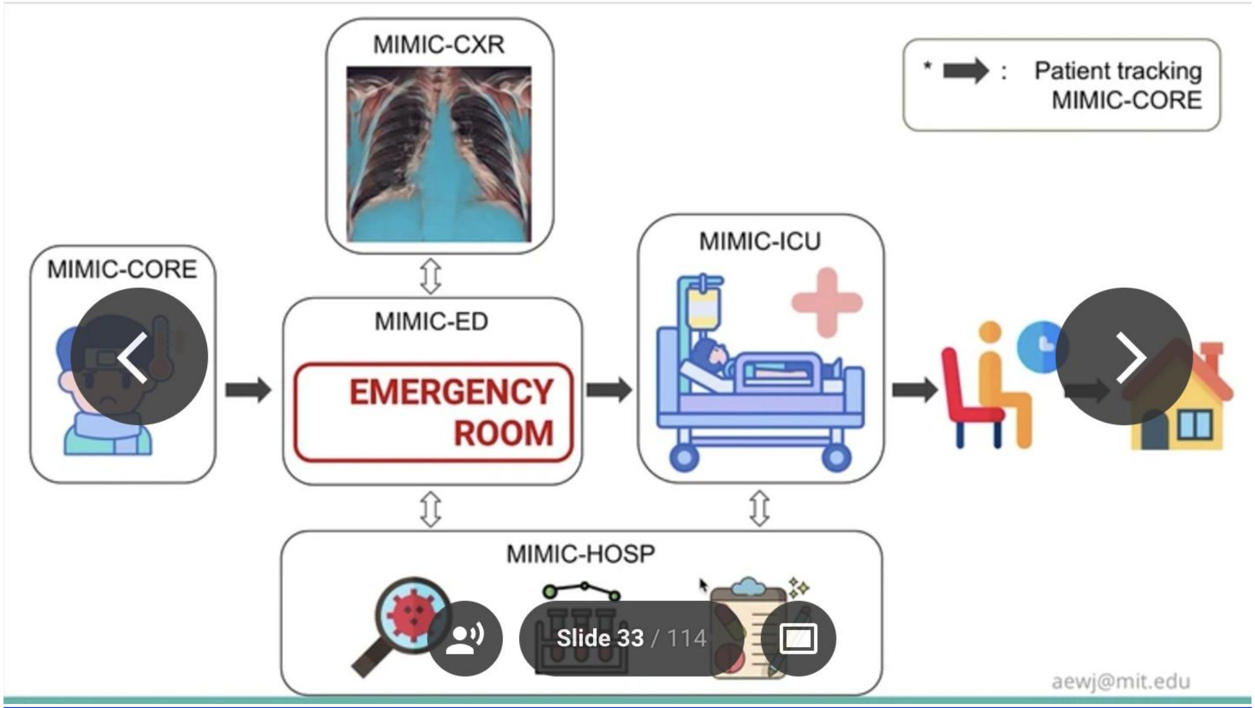
Results: Fig *** indicates prediction of the DL Model for LVEF in (a) a *** year old man, (b) ** year old woman using the DL based 12-lead ECG and single V2-V3 ECG lead. overall, the 12-lead ECG provided an AUC for detecting drop in LVEF of 0.88. Using a single V2-V3 bipole to reconstruct the 12-lead ECG approached provided an AUC of 0.77 to detect a drop in LVEF, while the single-lead V2-V3 alone provided AUC=0.60 of **. There was no improvement when adding clinical or demographic features to the model.

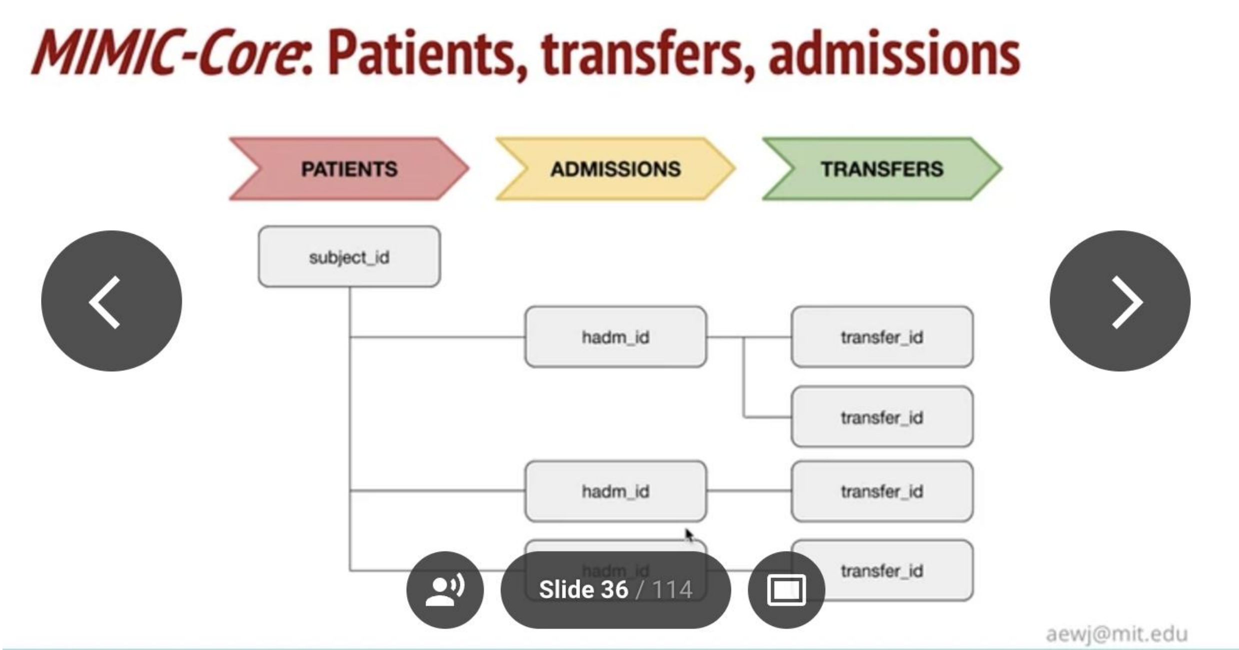
Conclusion: We show that deep convolutional neural networks applied to a single precordial ECG lead, such as from routine ambulatory recorders, can detect longitudinal trends in echocardiographic left ventricular ejection fraction. This approach opens avenues to inexpensive, continuous monitoring of physiological parameters that could be of use in rural, remote and diverse patient populations.

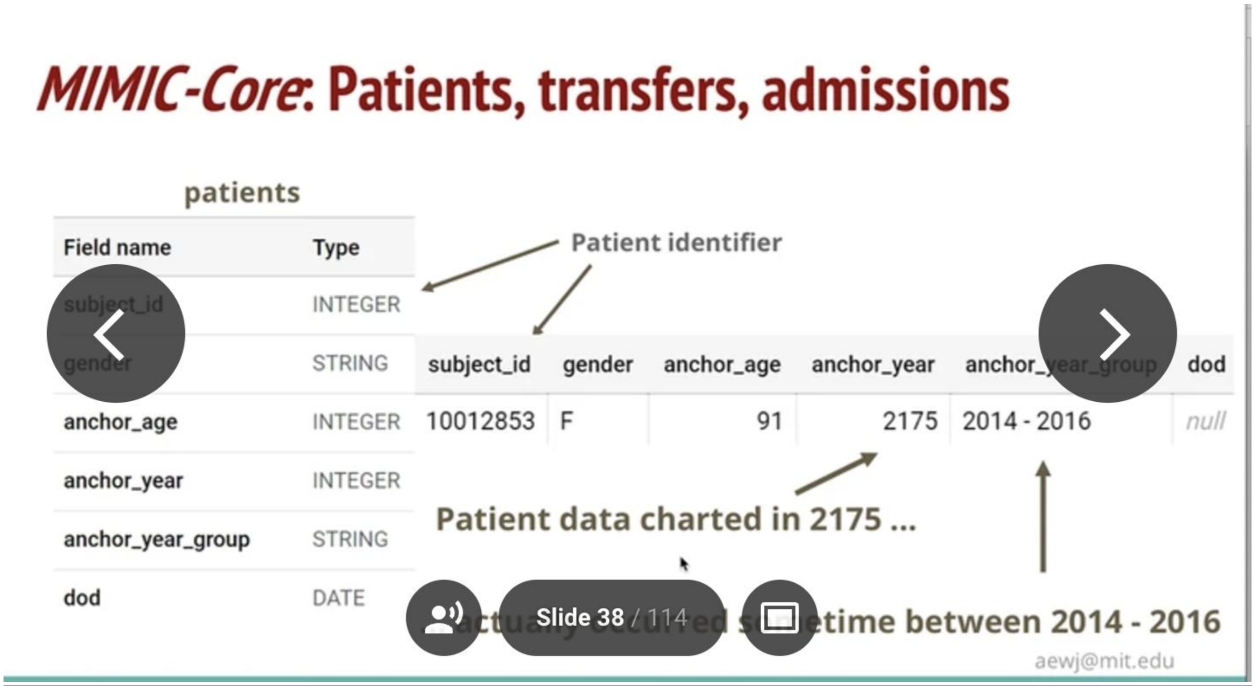
See video on using MIMIC
https://slideslive.com/embed/presentation/38931965



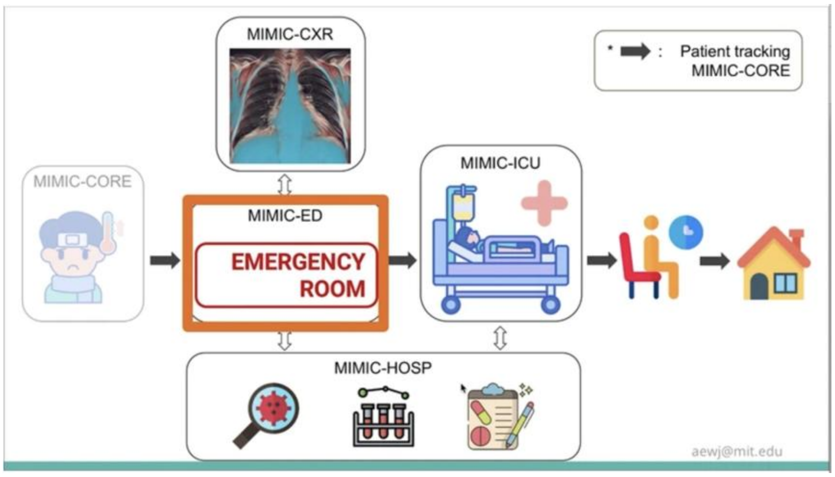


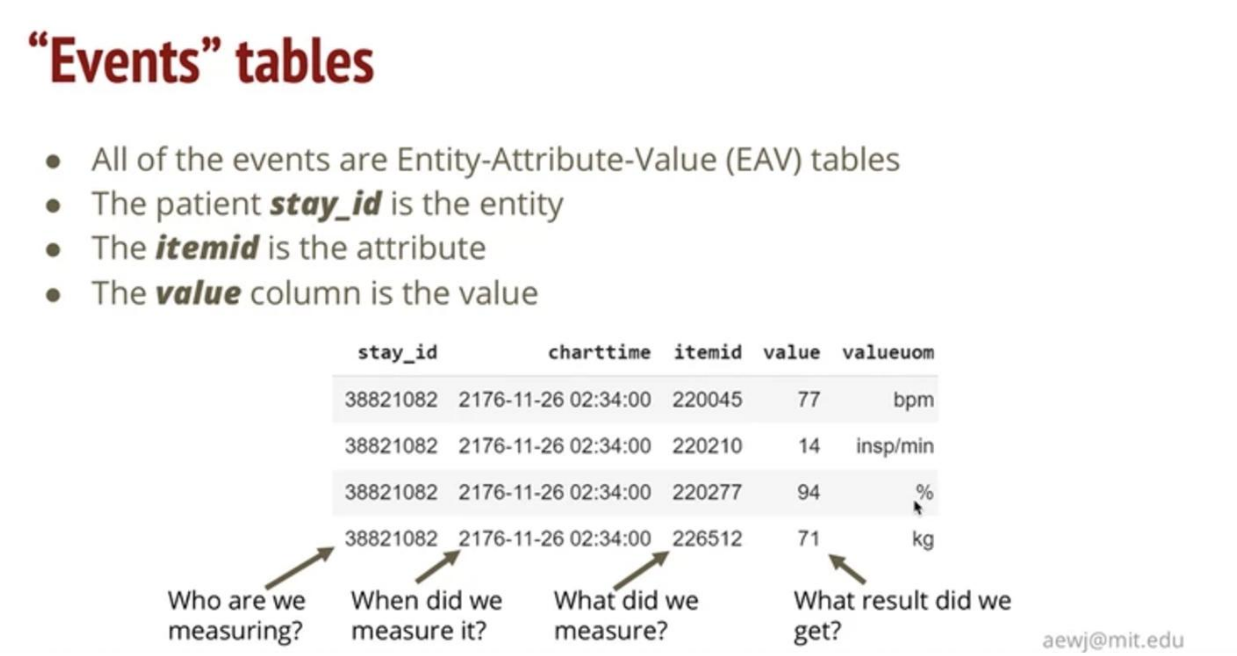

MIMIC-CXR
* ➡ : Patient tracking MIMIC-CORE
MIMIC-CORE
MIMIC-ED
EMERGENCY ROOM
MIMIC-ICU
MIMIC-HOSP
Slide 33 / 114
aewj@mit.edu


MIMIC-Core: Patients, transfers, admissions
PATIENTS
ADMISSIONS
TRANSFERS
subject_id
hadm_id
transfer_id
transfer_id
hadm_id
transfer_id
transfer_id
Slide 36 / 114
aewj@mit.edu


MIMIC-Core: Patients, transfers, admissions
patients
Field name
Type
subject_id
INTEGER
gender
STRING
anchor_age
INTEGER
anchor_year
INTEGER
anchor_year_group
STRING
dod
DATE
Patient identifier
subject_id
gender
anchor_age
anchor_year
anchor_year_group
dod
10012853
F
91
2175
2014 - 2016
null
Patient data charted in 2175 ...
time between 2014 - 2016
aewj@mit.edu
Slide 38 / 114

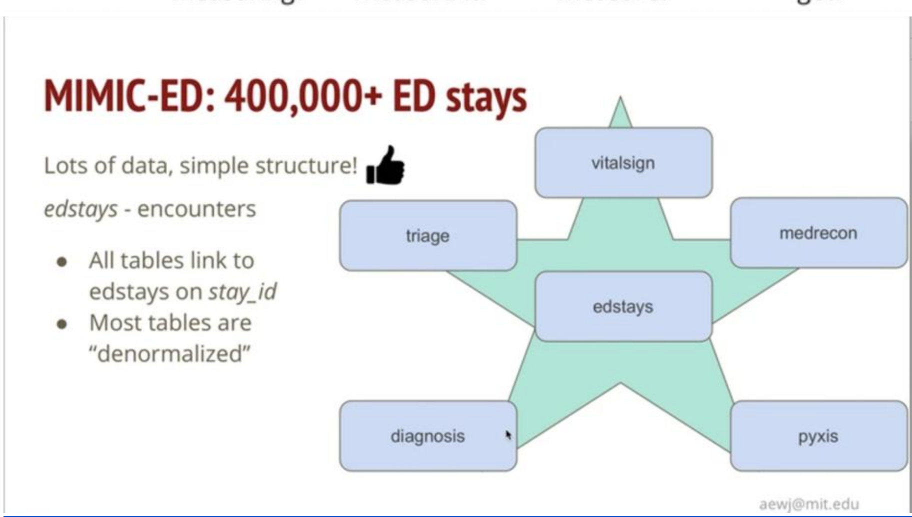
Recent data


* ➡ : Patient tracking
MIMIC-CORE
MIMIC-CXR
MIMIC-CORE
MIMIC-ED
EMERGENCY ROOM
MIMIC-ICU
MIMIC-HOSP
aewj@mit.edu


"Events" tables
All of the events are Entity-Attribute-Value (EAV) tables
The patient stay_id is the entity
The itemid is the attribute
The value column is the value

stay_id	charttime	itemid	value	valueuom
38821082	2176-11-26 02:34:00	220045	77	bpm
38821082	2176-11-26 02:34:00	220210	14	insp/min
38821082	2176-11-26 02:34:00	220277	94	%
38821082	2176-11-26 02:34:00	226512	71	kg

Who are we measuring?
When did we measure it?
What did we measure?
What result did we get?
aewj@mit.edu


MIMIC-ED: 400,000+ ED stays
Lots of data, simple structure!
edstays - encounters
All tables link to edstays on stay_id
Most tables are "denormalized"
vitalsign
triage
medrecon
edstays
diagnosis
pyxis
aewj@mit.edu

MIMIC-ED: *triage*

Assessment when the patient arrives to the ED ("triage")

stay_id	temperature	heartrate	resprate	o2sat	sbp	dbp	pain	acuity	chiefcomplaint
30056787	97.6	72	16	97	173	60	0	2	DVT, Transfer

MIMIC-ED: *vitalsign*

*P*eriodic vital signs recorded while the patient is in the ED

subject_id	stay_id	charttime	temperature	heartrate	resprate	o2sat	sbp	dbp	rhythm	pain
10012853	30056787	2180-07-24 18:37:00	97.6	72	16	97	173.0	60.0	None	0
10012853	30056787	2180-07-24 21:00:00	None	65	21	96	NaN	NaN	None	None
10012853	30056787	2180-07-24 21:34:00	None	69	21	96	164.0	51.0	None	0
10012853	30056787	2180-07-24 21:59:00	98	63	24	94	155.0	50.0	None	0
10012853	30056787	2180-07-24 22:46:00	None	71	15	93	163.0	59.0	None	0
10012853	30056787	2180-07-24 23:24:00	98.4	77	25	97	164.0	48.0	None	0
10012853	30056787	2180-07-24 23:24:00	98.4	77	25	97	164.0	48.0	None	0

aewj@mit.edu

MIMIC-ED: *medrecon*

Medicine reconciliation, i.e. what were you taking at home?

subject_id	stay_id	charttime	name	gsn	ndc	etc_rn	etccode	etcdescription
10012853	30056787	2180-07-24 19:42:00	amiodarone	017241	16714084301	1	00002734	Antiarrhythmic - Class III
10012853	30056787	2180-07-24 19:47:00	methimazole	006675	11528030001	1	00006080	Antithyroid Agents, Thionamides - Imidazole De...
10012853	30056787	2180-07-24 19:48:00	calcium carbonate-vit D3-min	062773	11822489880	1	00000734	Minerals and Electrolytes - Calcium Replacemen...
10012853	30056787	2180-07-24 19:45:00	lisinopril	041567	16729019801	1	00000224	ACE Inhibitors
10012853	30056787	2180-07-24 19:47:00	ferrous sulfate	062753	54868632400	1	00000785	Minerals and Electrolytes - Iron
10012853	30056787	2180-07-24 19:46:00	Coumadin	018080	15330026601	1	00000806	Anticoagulants - Coumarin

aewj@mit.edu

MIMIC-ED: *diagnosis*

Diagnosis billed on **discharge** from the ED

subject_id	stay_id	seq_num	icd_code	icd_title
10012853	30056787	1	I82.622	Acute embolism and thrombosis of deep veins of...

aewj@mit.edu

“Events” tables

- All of the events are Entity-Attribute-Value (EAV) tables
- The patient ***stay_id*** is the entity
- The ***itemid*** is the attribute
- The ***value*** column is the value

stay_id	charttime	itemid	value	valueuom
38821082	2176-11-26 02:34:00	220045	77	bpm
38821082	2176-11-26 02:34:00	220210	14	insp/min
38821082	2176-11-26 02:34:00	220277	94	%
38821082	2176-11-26 02:34:00	226512	71	kg

Who are we measuring?

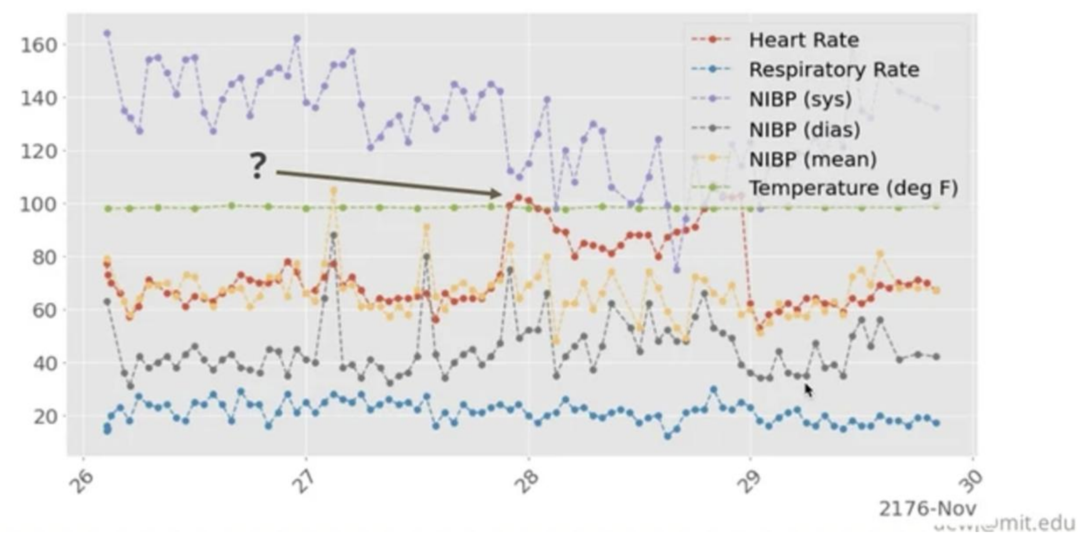
When did we measure it?

What did we measure?

What result did we get?

aewj@mit.edu

Vital signs for this patient



aewj@mit.edu

datetimeevents, outputevents

Similar to *chartevents*, but the data type of the value column is different

Numeric

outputevents

subject_id	hadm_id	stay_id	charttime	storetime	itemid	value	valueuom
18106347	24305596	32521459	2113-09-09 12:30:00	2113-09-09 13:38:00	226627	670.0	ml
18106347	24305596	32521459	2113-09-09 12:45:00	2113-09-09 13:37:00	226588	0.0	ml

datetimeevents

subject_id	hadm_id	stay_id	charttime	storetime	itemid	value
18106347	24305596	32521459	2113-09-09 12:40:00	2113-09-09 12:44:00	224292	2110-01-11 00:00:00
18106347	24305596	32521459	2113-09-09 12:40:00	2113-09-09 12:44:00	224298	2110-01-11 00:00:00

Datetime

aewj@mit.edu

procedureevents: events over time

charttime is replaced with *starttime* and *endtime*

subject_id	hadm_id	stay_id	starttime	endtime	storetime	itemid	value	valueuom
13859862	25015072	35387083	2155-12-05 01:50:00	2155-12-08 18:57:00	2155-12-08 19:07:00.000	225792	89.116667	hour
18917458	28038802	31785923	2189-08-13 21:11:00	2189-08-18 09:59:00	2189-08-19 08:31:00.000	225792	108.800000	hour

inputevents: inputs received over time

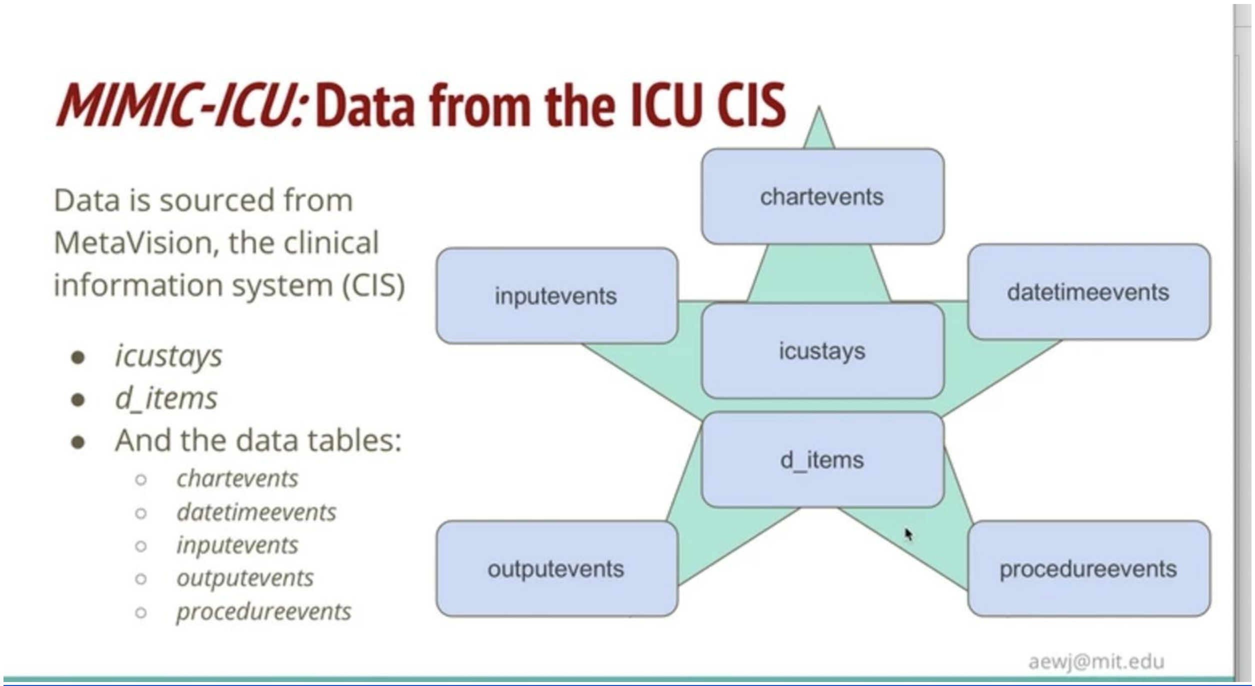
Has a *rate* and an *amount* column

subject_id	hadm_id	stay_id	starttime	endtime	storetime	itemid	amount	amountuom	rate	rateuom
10012853	27882036	38821082	2176-11-26 02:37:00	2176-11-26 12:27:00	2176-11-26 02:38:00	220949	113.113495	ml	11.503067	mL/hour
10012853	27882036	38821082	2176-11-26 02:37:00	2176-11-26 12:27:00	2176-11-26 02:38:00	225152	11311.349983	units	1150.306763	units/hour

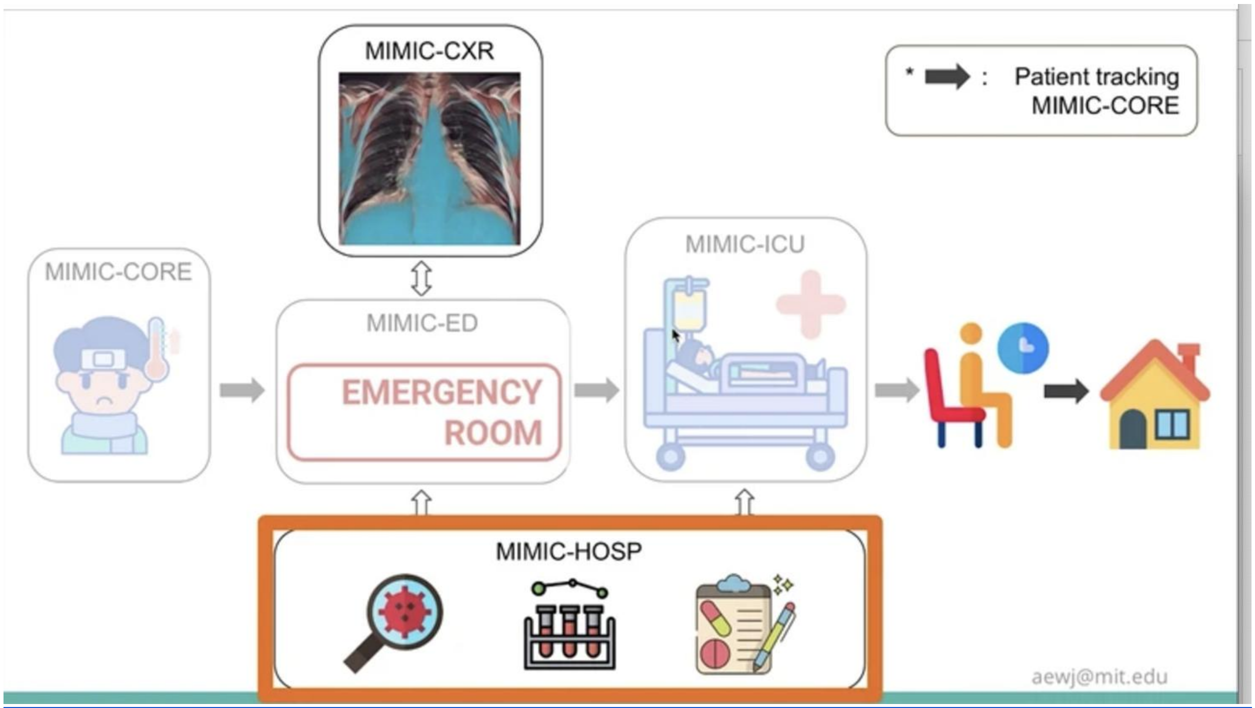
9 hours, 50 minutes

11311 units total == 1150 units/hour

aewj@mit.edu


MIMIC-ICU: Data from the ICU CIS
Data is sourced from MetaVision, the clinical information system (CIS)
icustays
d_items
And the data tables:
chartevents
datetimeevents
inputevents
outputevents
procedureevents
chartevents
inputevents
datetimeevents
icustays
d_items
outputevents
procedureevents
aewj@mit.edu

MIMIC – HOSPITAL
Comes from Hospital EHR



Note – designed for billing so may not comprehensively cover all events

MIMIC-Hosp: Billing data

- diagnoses_icd, d_icd_diagnoses
  - The diagnoses billed for in association with the patient's hospital visit
- drgcodes
  - Classifies the hospital into a diagnosis related group
- procedures_icd, d_icd_procedures
  - Billed procedures occurring during the patient's hospitalization
- hcpcsevents, d_hcpcs
  - Contains CPT codes and other billed healthcare events

aewj@mit.edu

MIMIC-Hosp: Billing data

Procedures billed by the hospital

- procedures_icd

Only one procedure for this patient

... though probably more than one thing happened during their hospital stay.

subject_id	hadm_id	seq_num	icd_code	icd_version	long_title
10012853	27882036	1	5A09557	10	Assistance with Respiratory Ventilation, Great...

Assistance with Respiratory Ventilation,
Greater than 96 Consecutive Hours,
Continuous Positive Airway Pressure

aewj@mit.edu

Ontology for MIMIC-Hosp not available

MIMIC-Hosp: Billing data

- *hcpcsevents*

subject_id	hadm_id	hcpcs_cd	seq_num	short_description

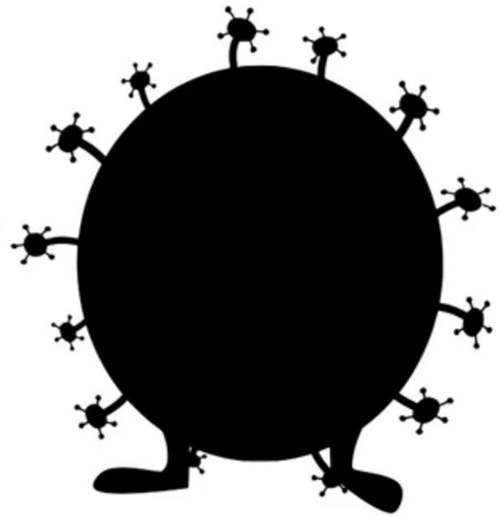
- No data for this patient
  - Not all tables will have data for all patients!
- All rows have an hcpcs_cd
- … but the meaning of the code costs $$$
  - Not a very useful table anyway!

MIMIC-Hosp: Microbiology

subject_id	charttime	storetime	test_name
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE
The patient	When the sample was taken	When the result came back	The test performed

Microbiology **takes time!**

The result is **not available** when the sample is taken



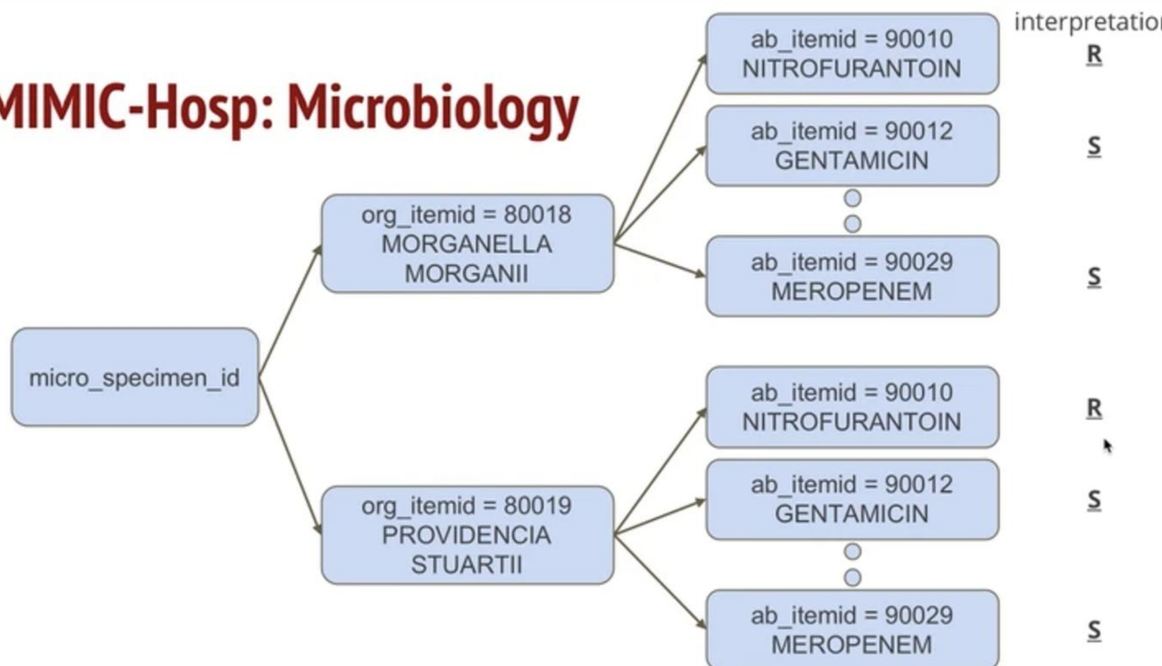
aewj@mit.edu

MIMIC-Hosp: Microbiology

subject_id	charttime	storetime	test_name	org_label	ab_label	dilution_text	interpretation
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	TRIMETHOPRIM/SULFA	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	NITROFURANTOIN	128	R
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	GENTAMICIN	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	TOBRAMYCIN	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	CEFTAZIDIME	None	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	CEFTRIAXONE	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	CIPROFLOXACIN	<=0.25	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	PIPERACILLIN/TAZO	<=4	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	CEFEPIME	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	MORGANELLA MORGANII	MEROPENEM	<=0.25	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	PROVIDENCIA STUARTII	TRIMETHOPRIM/SULFA	<=1	S
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	PROVIDENCIA STUARTII	NITROFURANTOIN	256	R
10012853	2176-10-06 02:37:00	2176-10-10 14:17:00	URINE CULTURE	PROVIDENCIA STUARTII	GENTAMICIN	<=1	S

aewj@mit.edu

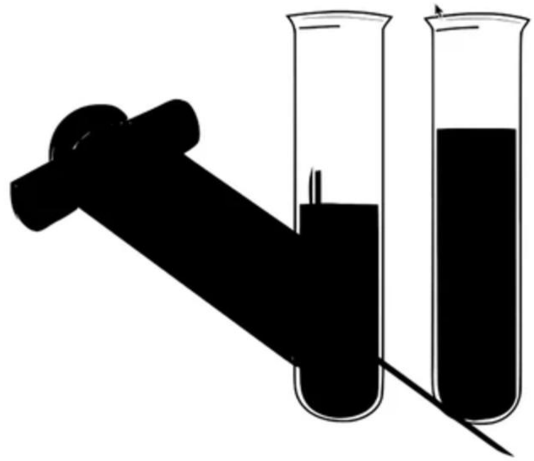
MIMIC-Hosp: Microbiology



aewj@mit.edu

MIMIC-Hosp: Labs

- Specimens drawn from a patient
- Measurements made on the specimen
    - Blood sample -> electrolytes
    - Urine sample -> pH
    - Cerebrospinal fluid -> white blood cell count



aewj@mit.edu

MIMIC-Hosp: Data sourced from the hospital EHR

Billing

hcpcsevents
d_hcpcs

diagnoses_icd
d_icd_diagnoses

procedures_icd
d_icd_procedures

drgcodes

Measurements

microbiologyevents
d_micro

labevents
d_labitems

Medications

prescriptions
pharmacy

emar
emar_detail

Admin

poe
poe_detail

services

aewj@mit.edu

prescriptions, pharmacy, emar, emar_detail

- prescriptions
  - Medications ordered for a patient
- pharmacy
  - Details about those medications from the pharmacy
- emar, emar_detail
  - Electronic medicine administration record
  - Digitally captures all administrations of a compound

aewj@mit.edu

MIMIC-Hosp: Administrative data

Information from *poe_detail*

subject_id	poe_id	ordertime	order_type	order_subtype	field_name	field_value
10012853	10012853-203	2176-11-26 00:09:39	ADT orders	Admit	Admit category	Place in observation
10012853	10012853-209	2176-11-26 01:32:13	ADT orders	Admit	Admit category	Admit to inpatient
10012853	10012853-215	2176-11-26 01:32:13	General Care	Tubes/Drains	Tubes & Drains type	Indwelling urinary catheter (IUC) - Foley
10012853	10012853-223	2176-11-26 01:45:45	General Care	Code status	Code status	Resuscitate (Full code)
10012853	10012853-225	2176-11-26 03:09:58	ADT orders	Transfer	Transfer to	MICU - Orange
10012853	10012853-225	2176-11-26 03:09:58	ADT orders	Transfer	Admit category	Admit to inpatient
10012853	10012853-232	2176-11-26 03:09:58	General Care	Code status	Code status	Resuscitate (Full code)
10012853	10012853-295	2176-11-27 12:39:20	IV therapy	IV access request	Indication	Hydration
10012853	10012853-337	2176-11-29 21:55:48	ADT orders	Transfer	Admit category	Admit to inpatient
10012853	10012853-359	2176-11-29 21:55:48	General Care	Code status	Code status	Resuscitate (Full code)
10012853	10012853-365	2176-11-29 21:55:48	General Care	Tubes/Drains	Tubes & Drains type	Indwelling urinary catheter (IUC) - Foley
10012853	10012853-377	2176-12-01 10:06:35	General Care	Tubes/Drains	Tubes & Drains type	Indwelling urinary catheter (IUC) - Foley

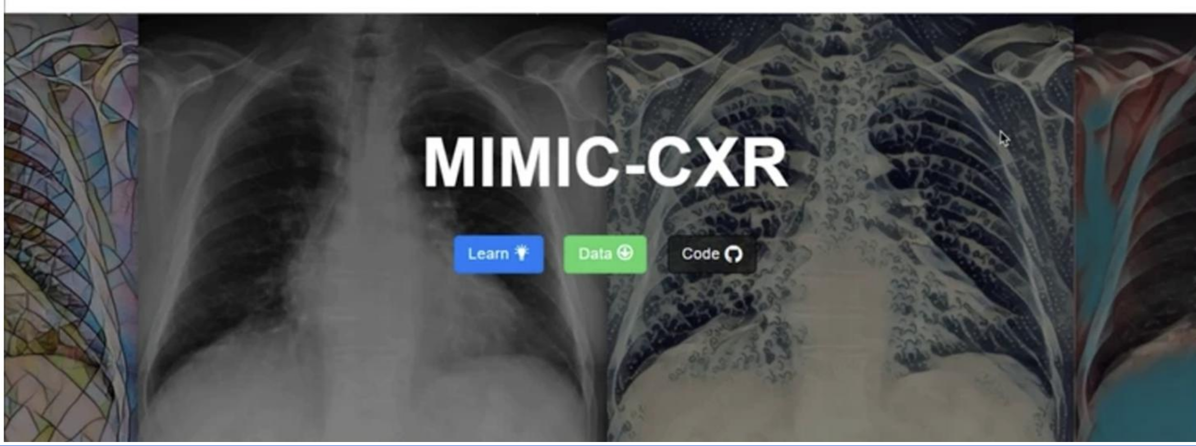
Our patient's story

- Admitted via the emergency department to the trauma/surgical ICU
- Transferred from the ED to the hospital for an acute embolism
- Had paroxysmal atrial fibrillation in the ICU
- Received non-invasive respiratory support for 96+ hours
- Blood gas monitoring shows oxygenation issues
- Blood clotting time was in a therapeutic range, treated with heparin IV
- Had chest x-rays ordered
- Was full code (patient wants to be resuscitated if necessary)

aewj@mit.edu

MIMIC-CXR: Chest x-rays for hospitalized patients

https://mimic-cxr.mit.edu



MIMIC-CXR + MIMIC-IV

If we look in MIMIC-CXR for this patient, we turn up 6 x-rays!
Two of these occur during the same hospitalization (the last two).

dicom	StudyID	PatientID	StudyDate	StudyTime	BodyPartExamined	ViewPosition
f6a37fec-15ac1f40-bf0526fc-39db0719-2b908d7a	53243235	10012853	21750405	062844.703	PORT CHEST	AP
aac02704-c647b84c-58d5df12-c9857852-e1536bba	58181999	10012853	21750407	131841.031	CHEST	AP
da11ee9c-deaeb30c-7017c575-d04ac857-bf990bc1	58569460	10012853	21760606	144043.421	CHEST	LATERAL
7dcc1738-1674a7a7-42bb1860-d1551597-33f6de53	58569460	10012853	21760606	144043.421	CHEST	PA
4736face-6f5f0e13-4b8bac9f-49661d81-a988229c	50200959	10012853	21761128	150953.187	CHEST	AP
683930ea-a24d96eb-f0ef3736-3906153d-0ac7d866	50200959	10012853	21761128	150953.187	CHEST	AP

In fact, recall there was an order in *poe* for the x-ray:

poe_id	poe_seq	subject_id	hadm_id	ordertime	order_type	order_subtype
10012853-320	320	10012853	27882036	2176-11-28 14:40:28	Radiology	General Xray

MIMIC-CXR + MIMIC-IV - The x-rays

Note the top right corner has been deidentified

IMPRESSION:

1. Moderate pulmonary congestion and mild interstitial edema is increased, moderate right pleural effusion is new, and moderate left basilar atelectasis is increased since ___, consistent with acute CHF exacerbation.
2. Large goiter.

Our patient's story

- Admitted via the emergency department to the trauma/surgical ICU
- Transferred from the ED to the hospital for an acute embolism
- Had paroxysmal atrial fibrillation in the ICU
- Received non-invasive respiratory support for 96+ hours
- Blood gas monitoring shows oxygenation issues
- Blood clotting time was in a therapeutic range, treated with heparin IV
- Had chest x-rays ordered
- Was full code (patient wants to be resuscitated if necessary)
- Signs of acute exacerbation of chronic heart failure on CXR
- Goiter present on CXR

The actual story (from the note!)

```
Brief Hospital Course:
___ y/o F with atrial fibrillation on warfarin, PE, CKD III,
PVD, multinodular goiter s/p biopsy w/ possible follicular
neoplasm in ___ who was sent to ED after being found to have
L jugular and subclavian venous thrombosis despite
therapeutic INR on warfarin.
...
#Hypoxemia and Hypercapnic Respiratory Failure ...
#LUE DVTs ...
#Gout ...
#Atrial Fibrillation ...
```