The Story of MIMIC II: Past, Present & Future
The creation of new knowledge and new clinical tools is uniquely dependent on the availability of relevant DATA.

Progress is accelerated by OPEN DATA, and is stifled by barriers to data sharing.
My Personal Time Series

- 1980’s  Case Study: Arrhythmia analysis
- 1999  PhysioNet launched
- 2000’s  MIMIC II
- > 2014  MIMIC III ????
Real-time Automated Arrhythmia Analysis
From Cottage Industry to Universal Standards

- Arrhythmia detector research 1960’s-70’s – a “cottage industry”
  - Proprietary data for algorithm development and evaluation

- Publicly distributed data
  - MIT-BIH Arrhythmia Database released publicly in 1980, 100,000 annotated beats from 48 Holter records
  - AHA Ventricular Arrhythmia Database – mid 1980s

- Open data:
  - Facilitated algorithm evaluation
  - Encouraged competition
  - Resulted in standardized metrics
  - Accelerated development of arrhythmia

- Value of open source data established
PhysioNet Established September, 1999

PhysioNet
the research resource
for complex
physiologic signals

Supported by
NIBIB and NIGMS
What is PhysioNet?
A unique web-based resource designed to support current research and stimulate new investigations in the study of complex biomedical and physiologic signals.

Four closely interdependent components:

- Data repository (PhysioBank)
  - Makes large collections of physiologic and clinical data available in open, Internet accessible, archives

- Library of related software (PhysioToolkit)
  - Open-source software for exploration and analysis of physiologic data

- Collaborative secure workspace (PhysioNetWorks)

- Free-access website (physionet.org)

www.physionet.org
Impact of PhysioNet

- Open access to:
  - > 4 TB of data in 51 collections, ~ 35,000 recordings
- Data and software contributed by a wide range of investigators worldwide
- Users:
  - Over 40,000 from all 50 states and more than 140 nations.
- Visits
  - ~ 4,000 per day
- Downloads:
  - ~ 290 gigabytes of data per day.
- Publications:
  - ~ 100 scholarly publications per month
MIMIC I
(Multi-parameter Intelligent Monitoring for Intensive Care) Database

- 1992 – 1999
- Data acquisition from individual patients
- IRB required patient/HCP consent
- Special-purpose archiver via RS-232 connection
- High quality data (500 s/s, 12 bit precision)
- 90 records, 40 hr. duration
- Clinical data via manual abstraction of paper records
NIH grant awarded in 2003:

“The objective of this Bioengineering Research Partnership is … to develop and evaluate advanced ICU patient monitoring systems that will substantially improve the efficiency, accuracy, and timeliness of clinical decision making in intensive care.”
GOALS

- Mirror the ICU environment in the laboratory
  - Waveforms, numerics, complete clinical data

- Support:
  - Development and evaluation of algorithms for advanced patient monitoring and decision support
  - Retrospective clinical research in critical care

- Make data public
  - Engage the creative energy of the world-wide research community
Partners

- **MIT faculty**
  - Mark, Szolovits, Verghese

- **Beth Israel Deaconess Medical Center**
  - Intensivists, IT department

- **Philips Healthcare**
Optimal Environment at BIDMC

- Well established high quality hospital information system built over 35 years
- IT department and chief (John Halamka) very supportive
- ICUs use Philips’ equipment:
  - Intellivue monitoring network
  - CareVue clinical information system
  - Information Storage Mart (ISM) – Oracle DBS
IRB Approval

- IRB approval 2001 and annually thereafter
- Waiver of informed consent
  - Data to be de-identified
  - Involves no risk to subjects
  - Waiver will not adversely affect the rights and welfare of the subjects
  - Research could not practically be carried out without the waiver
MIMIC II Data Acquisition

- Waveforms, numeric trends, alarms from bedside monitor network
- Archived clinical data from bedside workstation network
- Additional clinical data, discharge summaries, ICD-9 codes, POE, etc.
- Social Security Death Records

MIMIC II Waveform Database Clinical Database (Oracle)
Specific Tasks and Challenges

- **Waveforms**
  - Special purpose archiving agents supplied by Philips (unsupported)
  - Convert proprietary data format to WFDB
  - Match to clinical data
  - De-identify and time shift

- **Clinical Data**
  - Collect from ISM & hospital data extracts
  - Integrate into Oracle relational DBS
  - De-identify and time shift
Clinical Data for entire ICU stay & subsequent hospital admissions

- Hourly physiologic measures
- Laboratory results
- Fluid balance
- IV medications
- Ventilator settings
- Demographics
- ICD-9 codes
- Physician orders
- Reports: radiology, echo, ECG
- Nurse progress notes
- Discharge summaries
- etc.
MIMIC II Contents (2)

- Continuous Physiologic Data

Waveforms (sampled @ 125 Hz)
10-bit resolution

Waveforms with IV Medication Record

Minute-by-minute trend of physiological measurements
Automated De-identification of Free Text

Individual’s Protected Health Information

De-identification

De-identified Patient Data
Protected Health Information (PHI)
Health Insurance Portability and Accountability Act
(HIPAA 2003)

- Names
- Geographic locations (smaller than a state)
- Dates (except year)
- Ages over 89
- Telephone/Pager #
- Fax #
- Email
- SSN
- Medical record #
- Health plan beneficiary #

- Certificate/license #
- Account #
- Vehicle IDs, serial #
- Device ID and serial #
- URLs
- IP address numbers
- Biometric identifiers (finger & voice prints)
- Full face photographic images
- Any other unique identifying number, characteristic, or code.
De-Identification Strategy

**Lexical Matching**

**PHI Lookup Tables**
- Patient names
- Doctor names
- Hospital names
- Company names
- First, Last names
- Locations

**Pattern Matching with Regular Expressions**

**Numeric Tokens**
- Dates, phone numbers
- SSN, zip codes, MRN

**Expressions with Indicators**
- More names and locations

**PHI Indicators**
- Name: Mr, MS, Dr, son, wife, ...
- Location: Street, Road, Town,
- Age: y.o., years old, yo ...

Mark words before or after Indicators as potential PHI

**Non-PHI**
- UMLS/ Snomed
- Common words

**Heuristics**
- Match with UMLS and common words
- Apply simple rules

hpi: Mr. Costanzo admitted in transfer from Calvert Hospital on 8/16/1995 following a clinic ...
. pt of dr. randy jones.

meds: pt claims that he is compliant .... , colace 100mg po qd, methadone (methadone in house has been dc'd) his wife, janice (508/334/5423) confirms the story.

hpi: Mr. Costanzo admitted in transfer from Calvert Hospital on 8/16/1995 following a clinic ...
. pt of dr. randy jones.

meds: pt claims that he is compliant .... , colace 100mg po qd, methadone (methadone in house has been dc'd) his wife, janice (508/334/5423) confirms the story.

hpi: [**Known patient name**] admitted in transfer from [**Hospital 1**] Hospital on [**2002 09-02**] following a clinic ... pt of dr. [**First Name 2**] [**Doctor Last Name 3**].

meds: pt claims that he is compliant ... colace 100mg po qd, methadone (methadone in house has been dc'd) his wife, [**Name2 4**] [**Telephone/Fax 5**] confirms the story.
## Performance on Reference Corpus

(350,000 words from 2,785 nursing progress notes)

<table>
<thead>
<tr>
<th></th>
<th>Recall (Sens.)</th>
<th>Precision (PPV)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single MD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td>0.81</td>
<td>0.98</td>
</tr>
<tr>
<td>Min</td>
<td>0.63</td>
<td>0.95</td>
</tr>
<tr>
<td>Max</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Team of two MDs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>Min</td>
<td>0.89</td>
<td>0.95</td>
</tr>
<tr>
<td>Max</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Team of three MD’s</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Min</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>Max</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Algorithm</strong></td>
<td>0.97</td>
<td>0.75</td>
</tr>
</tbody>
</table>

**FN rate** – 2 per 10,000 words  
No patient names missed

(BMC Medical Informatics and Decision Making 2008, 8:32)
hp: [**Known patient name**] admitted in transfer from [**Hospital 1**] Hospital on [**2012-09-02**] following a clinic visit where he c/o sob, ruq pain, and possible r/o chf. pt admitted to ccu on [**09-03**] for increased aggitation, twitching, periods of unresponsiveness, arf illustrated by a creatnine of 4 (1.2 1 month ago), and elevated liver enzymes. pt of dr. [**First Name 2**] [**Doctor Last Name 3**].

meds: pt claims that he is compliant at home on meds, amiodarone 200mg po qd, dig .125mg po qd, colace 100mg po qd, methadone (methadone in house has been dc'd) his wife, [**Name2 4**] [** Telephone/Fax 5**]) confirms the story.
MIMIC II – v2.6

  - 32,332 adults
  - 8,087 neonates

- Distribution (adults)
  - MICU - 40%
  - SICU – 25%
  - CSRU – 19%
  - CCU – 15%

- Hospital mortality (percent of unit stays) = 11.7%

- Records with matched waveforms & trends ~ 5,000

- Database size
  - Waveforms ~ 4 TB
  - Clinical ~ 200 GB
Version 3.0
Estimated release April 1, 2014

- Covers 2001 – 2012
- Adds 23,800 additional adult ICU admissions
- Incorporates new data from MetaVision
  - Physician notes
  - Medication administration
- Includes more complete hospital data
  - CPT codes
  - ADT files
  - Demographics/Admission data
- Reorganized, more denormalized table structure
“Open” MIMIC II

- Over 23,000 waveform records have been posted on PhysioNet - durations from days to several weeks. Free access to all.

http://physionet.org/physiobank/database/mimic2db/

- Full MIMIC II database is freely available, subject only to a data use agreement. Total current users ~ 950

www.physionet.org/cgi-bin/request

- I will not attempt to identify any individual in the DB
- I will not share access to data with anyone else
- I will maintain physical and electronic security of the data
- If I find any PHI I will report its location so it can be removed
- My purpose is lawful research
Signal Processing Studies

PubMed

Methods of blood pressure measurement in the ICU.
Lehman LW, Saeed M, Taimor D, Mark R, Malhotra A.

Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator.
Li Q, Mark RG, Clifford GD.

Signal quality estimation with multichannel adaptive filtering in intensive care settings.
Silva I, Lee J, Mark RG.

Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension.
Huq CW, Clifford GD, Reisner AT.

Accessing the public MIMIC-II intensive care relational database for clinical research.
Scott DJ, Lee J, Silva I, Park S, Moody GB, Celi LA, Mark RG.

The cardiac output from blood pressure algorithms trial.
Sun JX, Reisner AT, Saeed M, Heldt T, Mark RG.
Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform.
Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD.

Transfer entropy estimation and directional coupling change detection in biomedical time series.
Lee J, Nemati S, Shiva I, Edwards BA, Butler JP, Malhotra A.

Robust detection of premature ventricular contractions using a wave-based Bayesian framework.
Sayadi O, Shamsollahi MB, Clifford GD.

Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator.
Li Q, Mark RG, Clifford GD.

Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter.
Li Q, Mark RG, Clifford GD.

Continuous blood pressure-derived cardiac output monitoring--should we be thinking long term?
Heldt T.

Cycle-averaged dynamics of a periodically driven, closed-loop circulation model.
Heldt T¹, Chang JL, Chen JJ, Verghese GC, Mark RG.

Continuous cardiac output monitoring by peripheral blood pressure waveform analysis.
Mukkamala R, Reisner AT, Homan HM, Mark RG, Cohen RJ.
Clinical Studies

**Association of hypermagnesemia and blood pressure in the critically ill.**

**Severity of acute kidney injury and two-year outcomes in critically ill patients.**

**Proton-pump inhibitor use is associated with low serum magnesium concentrations.**
Danziger J, William JH, Scott DJ, Lee J, Lehman LW, Mark RG, Howell MD, Celi LA, Mukamal KJ.

**Leveraging a critical care database: SSRI use prior to ICU admission is associated with increased hospital mortality.**
Chassemi M, Marshall J, Singh N, Stone DJ, Celi LA.

**Disease-based Modeling to Predict Fluid Response in Intensive Care Units.**
Fialho AS, Celi LA, Cismondi F, Vieira SM, Reti SR, Sousa JM, Finkelstein SN.

**Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension.**
Maraud L, Lai PS, Clifford GD, Tarassenko L, Celi LA, Annane D.
Pharmacovigilance: an active surveillance system to proactively identify risks for adverse events.

Moses C, Celi LA, Marshall J.

Empirical relationships among oliguria, creatinine, mortality, and renal replacement therapy in the critically ill.

Mandelbaum T, Lee J, Scott DJ, Mark RG, Malhotra A, Howell MD, Talmor D.

A Database-driven Decision Support System: Customized Mortality Prediction.

Celi LA, Galvin S, Davidzon G, Lee J, Scott D, Mark R.

ICU admission characteristics and mortality rates among elderly and very elderly patients.

Fuchs L, Chronaki CE, Park S, Novack V, Baumfeld Y, Scott D, McNellen S, Talmor D, Celi L.

Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients.

Hunziker S, Celi LA, Lee J, Howell MD.

Interrogating a clinical database to study treatment of hypotension in the critically ill.

Lee J, Kothari R, Ladapo JA, Scott DJ, Celi LA.

A Clinical Database-Driven Approach to Decision Support: Predicting Mortality Among Patients with Acute Kidney Injury.

Celi LA, Tang R, Villarreal MC, Davidzon GA, Lester WT, Chuah HC.

Outcome of critically ill patients with acute kidney injury using the Acute Kidney Injury Network criteria.

Mandelbaum T, Scott DJ, Lee J, Mark RG, Malhotra A, Waikar SS, Howell MD, Talmor D.
Our Vision for MIMIC III

- Collaborate with colleagues to build a multi-institutional, international archive of highly granular critical care data that is open to all, and fully extensible.

- Generate new knowledge and new clinical tools to benefit patient care.

- Challenges to solve:
  - Technology development
  - Institutional support re data release
  - Long-term funding