“Machine Learning Amidst Health Record Data Irregularity: Subgrouping in Dimensions of Space and Time”

Jeremy C. Weiss

Assistant Professor of Health Informatics

Heinz College

Carnegie Mellon University

Real World Evidence: Machine Learning and Electronic Health Records

Jeremy C. Weiss, MD PhD
Assistant Professor of Health Informatics
Heinz College of Information Systems and Public Policy
Carnegie Mellon University

Why Real-World Evidence (RWE)?

Real-World Evidence — What Is It and What Can It Tell Us?

Rachel E. Sherman, M.D., M.P.H., Steven A. Anderson, Ph.D., M.P.P., Gerald J. Dal Pan, M.D., M.H.S., Gerry W. Gray, Ph.D., Thomas Gross, M.D., M.P.H., Nina L. Hunter, Ph.D., Lisa LaVange, Ph.D., Danica Marinac-Dabic, M.D., Ph.D., Peter W. Marks, M.D., Ph.D., Melissa A. Robb, B.S.N., M.S., Jeffrey Shuren, M.D., J.D., Robert Temple, M.D., Janet Woodcock, M.D., Lilly Q. Yue, Ph.D., and Robert M. Califf, M.D.

Trials are not enough:

- limited generalizability

- unknown interactions

- expensive

These trials are often needed because they are designed to provide an essential element of the premarket evaluation of a medical product — namely, robust evidence that a treatment may "work." However, the internal validity attained in these trials is often achieved at the expense of uncertainty about generalizability, especially since the populations enrolled in such studies may differ in significant ways from those seen in practice. In addition, there may be few data on interactions with concomitant illnesses and treatment, and adherence to therapies may be supported by intensive efforts that are infeasible in practice. Moreover, the expense of conducting large traditional trials has been growing steadily for years,[6] and recent estimates suggest that the cost trajectory may be steepening,[7] without any indication of a commensurate increase in the quantity of evidence produced to support decisions about health care.

The NEW ENGLAND JOURNAL of MEDICINE
December 8, 2016
N Engl J Med 2016; 375:2293-2297
DOI: 10.1056/NEJMsb1609216

What is Real-World Evidence (RWE)?

Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.

Real-World Evidence (RWE) is the *clinical* evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

Outline

I. Converting data to evidence: RWD to RWE
   Contrast with associative analysis best practices
   Machine learning tools for RWE (1) visualization and clustering

II. Machine learning tools for RWE (2) risk characterization

Part I: Visualizing and design choices for RWD

The promise of Real-World Data (RWD)

\+ Scale

\+ Representative populations

– Non-interventional means that causal questions are hard

– Uncontrolled

1. non-uniform data collection, follow-up
2. making the closed-world assumption

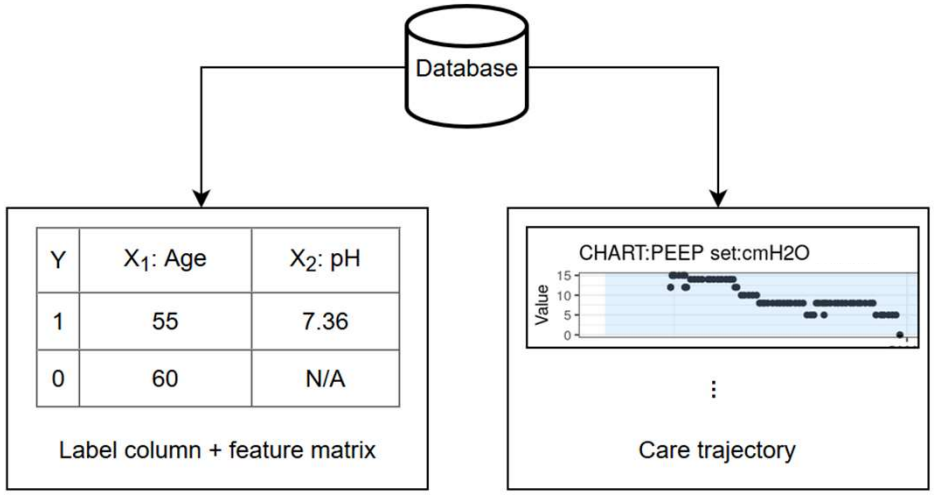
Evidence where trials are

Impractical
Infeasible
Impossible
Unethical

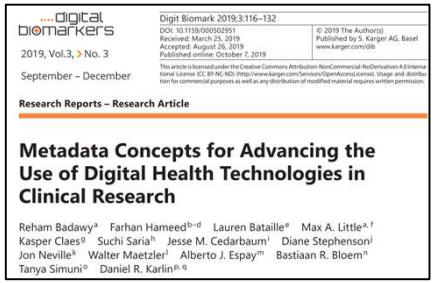
RWD and Machine Learning (ML)

Machine learning to augment the clinical process

- Clinical decision support: risk scores
- Public health: risk stratification
- Radiology and pathology: image segmentation/annotation



What should RWD include?



On the reliability of measurement

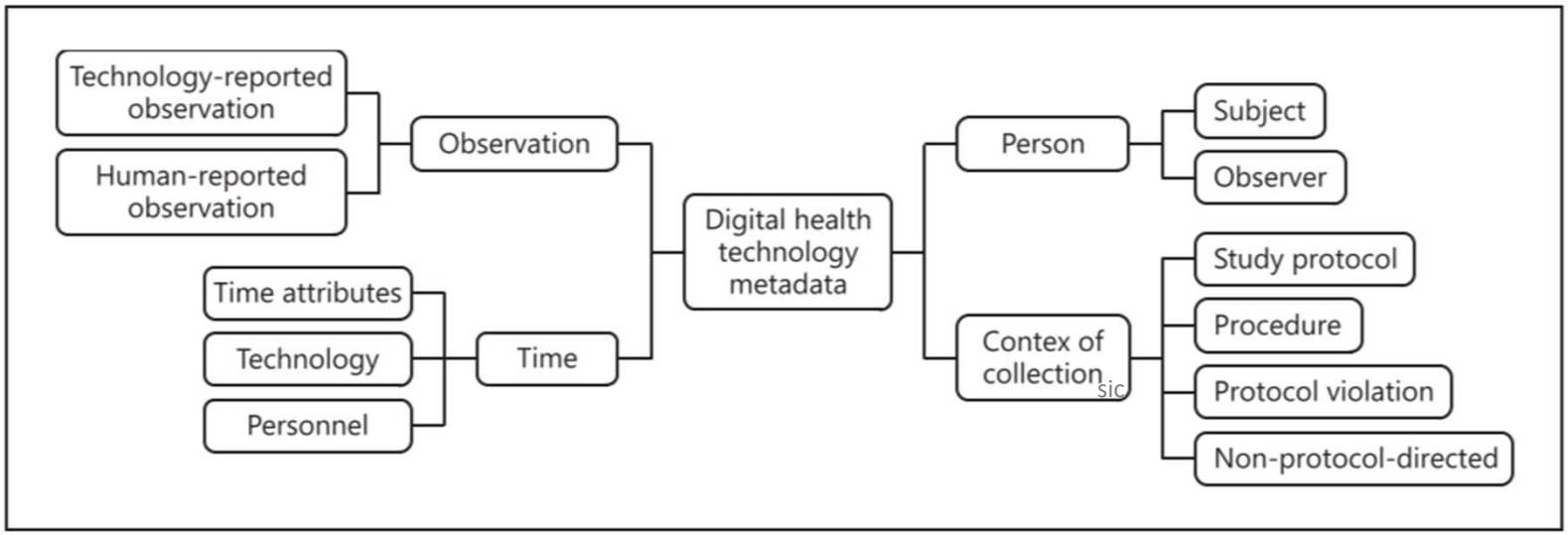
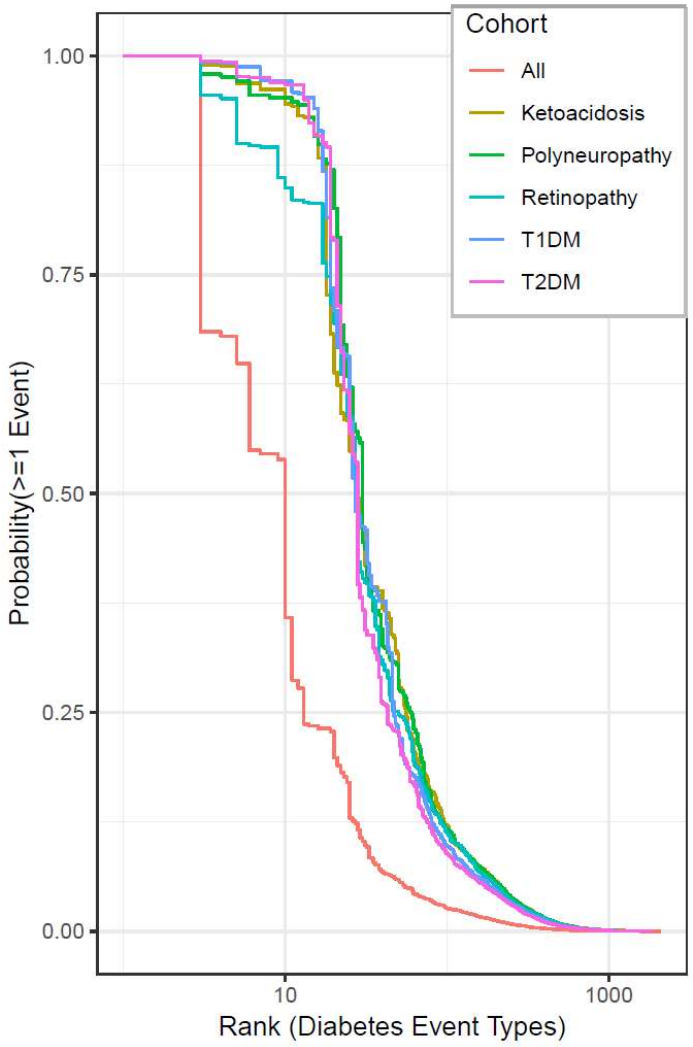


Fig. 1. Overview of the four fundamental concepts of the proposed digital health technology metadata set.

Not just outcomes (y) and features ($\boldsymbol{x}$), with text descriptions in the manuscript

Challenges of health records: non-uniform measurement collection



Data: Marshfield Clinic Electronic Health Record
Inclusion criteria:
Medical event 1960-2005, AND,
(medical event >2010, OR, death record >2005)
Population (n): 1.2 million patients

Choices:
- Find subgroups with low levels of missingness

- Analyze amidst missingness
How can we utilize the tail? With machine learning?

Find subgroups with low levels of missingness

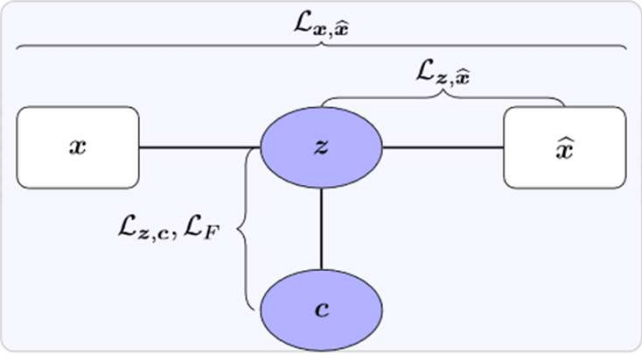
Simulations

Algorithmic translation: Clustering on counts or binary indicators

Problem: Majority of clustering methods rely heavily on distance measures (and usually Euclidean distance). Distance breaks down in high dimensions.

Revised problem: How to cluster high-dimensional sparse count data?

Solution: Hyperspherical clustering (Fillmore, Mehta, and Weiss, *AMIA Annual Symposium* 2019)



Background: variational autoencoders are used to disentangle latent features, assigning each an independent latent dimension of a multivariate normal (MVN) with k dimensions

Geometry: all density of a MVN is on the surface as $k \to \infty$

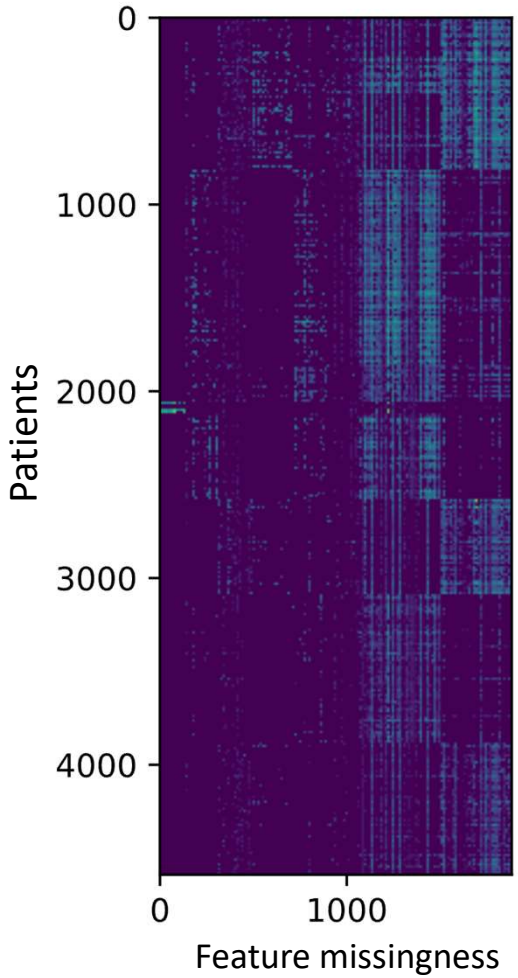
Result:

angle principally determines location, and,

angle admits a cluster probability: $[0.5, 0.866]^2 \rightarrow [0.25, 0.75]$

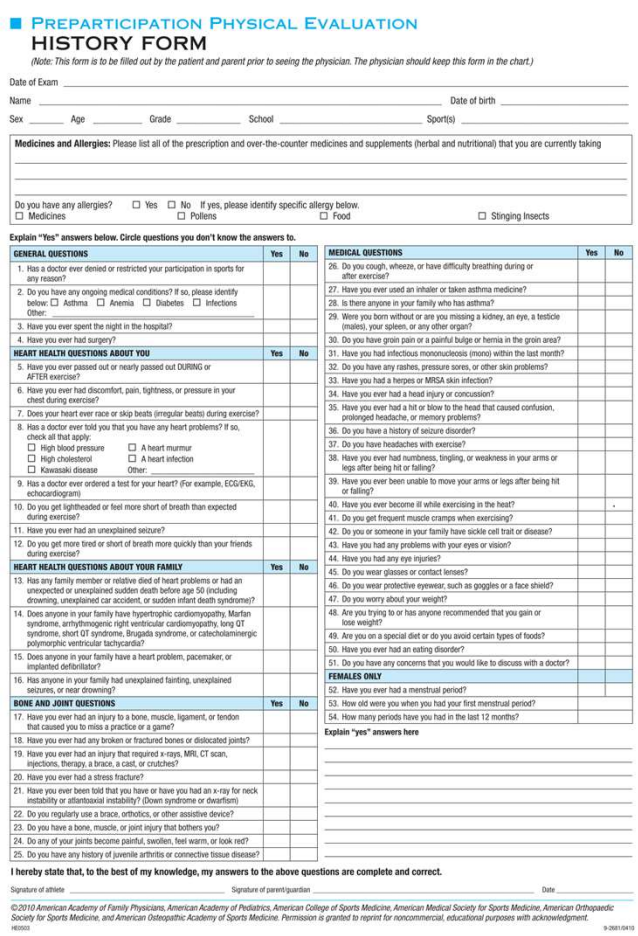
Find subgroups with low levels of missingness

Bicluster result on MIMIC III v1.4 dataset
Patients with ICD-9 code for pneumonia



Interpretation:
Multiple clusters of patients (7) and features (8)
Two primary sources of collected data:
- CareVue
- Metavision
In addition to these measurements are variably-measured features that cluster but demonstrate irregular patterns of collection

How well can we predict if we ignore the tail?



PREPARTICIPATION PHYSICAL EVALUATION
HISTORY FORM

(Note: This form is to be filled out by the patient and parent prior to seeing the physician. The physician should keep this form in the chart.)

Date of Exam _____

Name _____ Date of birth _____

Sex _____ Age _____ Grade _____ School _____ Sport(s) _____

Medicines and Allergies: Please list all of the prescription and over-the-counter medicines and supplements (herbal and nutritional) that you are currently taking

Do you have any allergies? ☐ Yes ☐ No If yes, please identify specific allergy below.
☐ Medicines ☐ Pollens ☐ Food ☐ Stinging Insects

Explain "Yes" answers below. Circle questions you don't know the answers to.

GENERAL QUESTIONS	Yes	No
1. Has a doctor ever denied or restricted your participation in sports for any reason?		
2. Do you have any ongoing medical conditions? If so, please identify below: ☐ Asthma ☐ Anemia ☐ Diabetes ☐ Infections Other:		
3. Have you ever spent the night in the hospital?		
4. Have you ever had surgery?		
HEART HEALTH QUESTIONS ABOUT YOU	**Yes**	**No**
5. Have you ever passed out or nearly passed out DURING or AFTER exercise?		
6. Have you ever had discomfort, pain, tightness, or pressure in your chest during exercise?		
7. Does your heart ever race or skip beats (irregular beats) during exercise?		
8. Has a doctor ever told you that you have any heart problems? If so, check all that apply: ☐ High blood pressure ☐ A heart murmur ☐ High cholesterol ☐ A heart infection ☐ Kawasaki disease Other:		
9. Has a doctor ever ordered a test for your heart? (For example, ECG/EKG, echocardiogram)		
10. Do you get lightheaded or feel more short of breath than expected during exercise?		
11. Have you ever had an unexplained seizure?		
12. Do you get more tired or short of breath more quickly than your friends during exercise?		
HEART HEALTH QUESTIONS ABOUT YOUR FAMILY	**Yes**	**No**
13. Has any family member or relative died of heart problems or had an unexpected or unexplained sudden death before age 50 (including drowning, unexplained car accident, or sudden infant death syndrome)?		
14. Does anyone in your family have hypertrophic cardiomyopathy, Marfan syndrome, arrhythmogenic right ventricular cardiomyopathy, long QT syndrome, short QT syndrome, Brugada syndrome, or catecholaminergic polymorphic ventricular tachycardia?		
15. Does anyone in your family have a heart problem, pacemaker, or implanted defibrillator?		
16. Has anyone in your family had unexplained fainting, unexplained seizures, or near drowning?		
BONE AND JOINT QUESTIONS	**Yes**	**No**
17. Have you ever had an injury to a bone, muscle, ligament, or tendon that caused you to miss a practice or a game?		
18. Have you ever had any broken or fractured bones or dislocated joints?		
19. Have you ever had an injury that required x-rays, MRI, CT scan, injections, therapy, a brace, a cast, or crutches?		
20. Have you ever had a stress fracture?		
21. Have you ever been told that you have or have you had an x-ray for neck instability or atlantoaxial instability? (Down syndrome or dwarfism)		
22. Do you regularly use a brace, orthotics, or other assistive device?		
23. Do you have a bone, muscle, or joint injury that bothers you?		
24. Do any of your joints become painful, swollen, feel warm, or look red?		
25. Do you have any history of juvenile arthritis or connective tissue disease?		

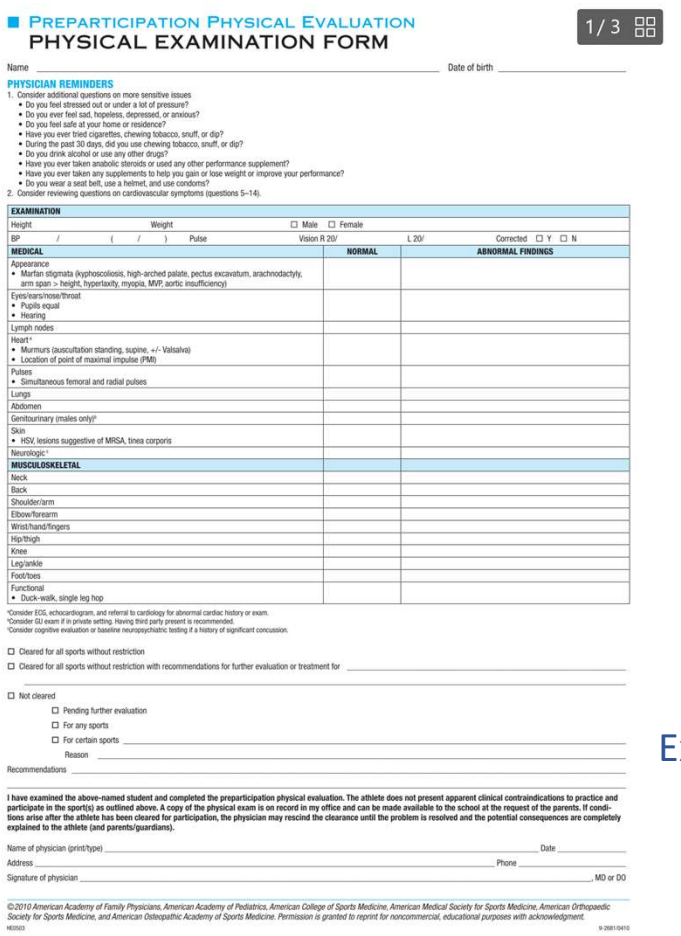
MEDICAL QUESTIONS	Yes	No
26. Do you cough, wheeze, or have difficulty breathing during or after exercise?		
27. Have you ever used an inhaler or taken asthma medicine?		
28. Is there anyone in your family who has asthma?		
29. Were you born without or are you missing a kidney, an eye, a testicle (males), your spleen, or any other organ?		
30. Do you have groin pain or a painful bulge or hernia in the groin area?		
31. Have you had infectious mononucleosis (mono) within the last month?		
32. Do you have any rashes, pressure sores, or other skin problems?		
33. Have you had a herpes or MRSA skin infection?		
34. Have you ever had a head injury or concussion?		
35. Have you ever had a hit or blow to the head that caused confusion, prolonged headache, or memory problems?		
36. Do you have a history of seizure disorder?		
37. Do you have headaches with exercise?		
38. Have you ever had numbness, tingling, or weakness in your arms or legs after being hit or falling?		
39. Have you ever been unable to move your arms or legs after being hit or falling?		
40. Have you ever become ill while exercising in the heat?		
41. Do you get frequent muscle cramps when exercising?		
42. Do you or someone in your family have sickle cell trait or disease?		
43. Have you had any problems with your eyes or vision?		
44. Have you had any eye injuries?		
45. Do you wear glasses or contact lenses?		
46. Do you wear protective eyewear, such as goggles or a face shield?		
47. Do you worry about your weight?		
48. Are you trying to or has anyone recommended that you gain or lose weight?		
49. Are you on a special diet or do you avoid certain types of foods?		
50. Have you ever had an eating disorder?		
51. Do you have any concerns that you would like to discuss with a doctor?		
FEMALES ONLY		
52. Have you ever had a menstrual period?		
53. How old were you when you had your first menstrual period?		
54. How many periods have you had in the last 12 months?		

Explain "yes" answers here

I hereby state that, to the best of my knowledge, my answers to the above questions are complete and correct.

Signature of athlete _____ Signature of parent/guardian _____ Date _____

HE0503 9-2681/0410



PREPARTICIPATION PHYSICAL EVALUATION
PHYSICAL EXAMINATION FORM

1 / 3

Name _____ Date of birth _____

PHYSICIAN REMINDERS

1. Consider additional questions on more sensitive issues
   - Do you feel stressed out or under a lot of pressure?
   - Do you ever feel sad, hopeless, depressed, or anxious?
   - Do you feel safe at your home or residence?
   - Have you ever tried cigarettes, chewing tobacco, snuff, or dip?
   - During the past 30 days, did you use chewing tobacco, snuff, or dip?
   - Do you drink alcohol or use any other drugs?
   - Have you ever taken anabolic steroids or used any other performance supplement?
   - Have you ever taken any supplements to help you gain or lose weight or improve your performance?
   - Do you wear a seat belt, use a helmet, and use condoms?
2. Consider reviewing questions on cardiovascular symptoms (questions 5–14).

EXAMINATION		
Height	Weight	☐ Male ☐ Female
BP / (/)	Pulse	Vision R 20/ L 20/ Corrected ☐ Y ☐ N
MEDICAL	**NORMAL**	**ABNORMAL FINDINGS**
Appearance • Marfan stigmata (kyphoscoliosis, high-arched palate, pectus excavatum, arachnodactyly, arm span > height, hyperlaxity, myopia, MVP, aortic insufficiency)		
Eyes/ears/nose/throat • Pupils equal • Hearing		
Lymph nodes		
Heart[a] • Murmurs (auscultation standing, supine, +/- Valsalva) • Location of point of maximal impulse (PMI)		
Pulses • Simultaneous femoral and radial pulses		
Lungs		
Abdomen		
Genitourinary (males only)[b]		
Skin • HSV, lesions suggestive of MRSA, tinea corporis		
Neurologic[c]		
MUSCULOSKELETAL		
Neck		
Back		
Shoulder/arm		
Elbow/forearm		
Wrist/hand/fingers		
Hip/thigh		
Knee		
Leg/ankle		
Foot/toes		
Functional • Duck-walk, single leg hop		

[a]Consider ECG, echocardiogram, and referral to cardiology for abnormal cardiac history or exam.
[b]Consider GU exam if in private setting. Having third party present is recommended.
[c]Consider cognitive evaluation or baseline neuropsychiatric testing if a history of significant concussion.

☐ Cleared for all sports without restriction

☐ Cleared for all sports without restriction with recommendations for further evaluation or treatment for _____

☐ Not cleared
- ☐ Pending further evaluation
- ☐ For any sports
- ☐ For certain sports _____
- Reason _____

Recommendations _____

I have examined the above-named student and completed the preparticipation physical evaluation. The athlete does not present apparent clinical contraindications to practice and participate in the sport(s) as outlined above. A copy of the physical exam is on record in my office and can be made available to the school at the request of the parents. If conditions arise after the athlete has been cleared for participation, the physician may rescind the clearance until the problem is resolved and the potential consequences are completely explained to the athlete (and parents/guardians).

Name of physician (print/type) _____ Date _____

Address _____ Phone _____

Signature of physician _____, MD or DO

HE0503 9-2681/0410

Name
CC
HPI
PMH
PSH
FH
SH
Meds/alls
Labs
ROS

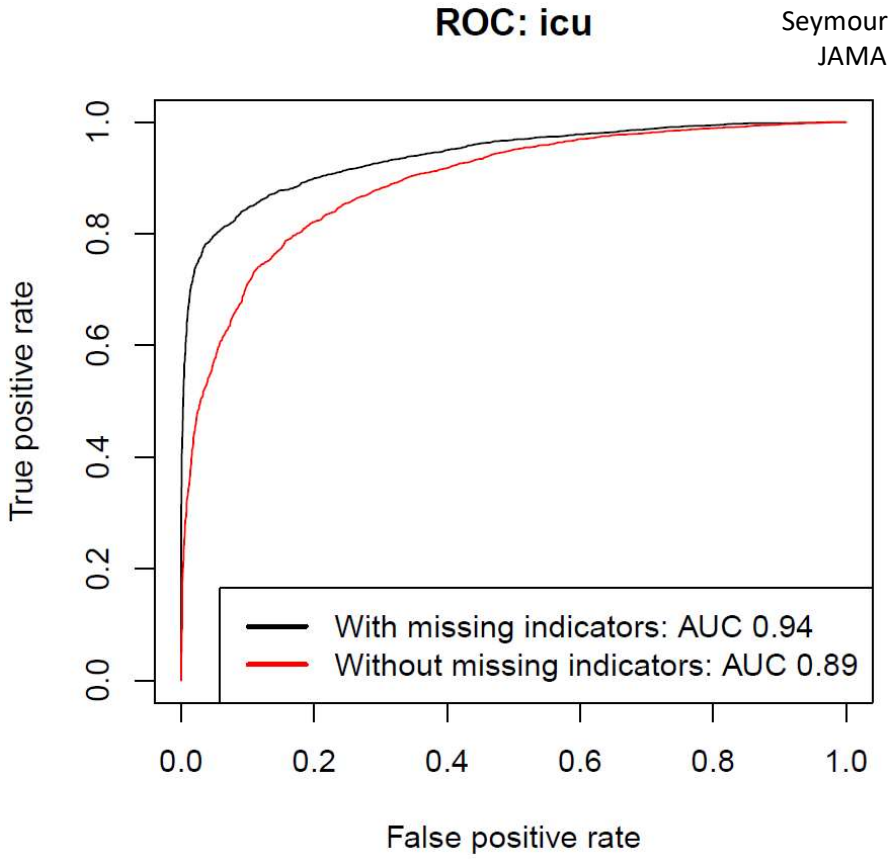
Exam / labs / imaging
Assessment
Plan

Either we limit our predictiveness and generalizability (narrow cohort), or we accept missingness.

The dilemma of missingness indicators

Data from Seymour et al, JAMA 2019

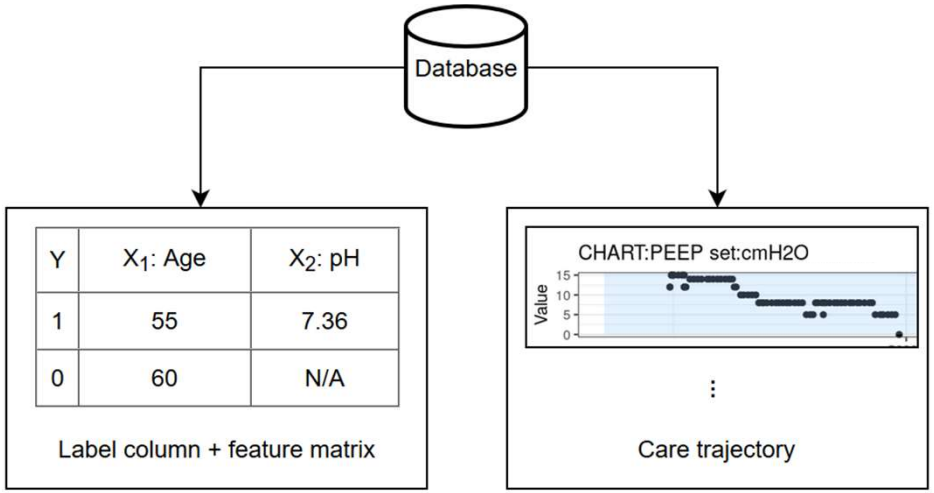
Survey: dyspnea (x)	Admit to ICU (y)	R[Survey]
N/A	1	1
1	0	0
0	0	0
0	0	0



Missingness indicators describe variation possibly distant from the patient: data collection protocol, hospital protocol, etc.

Missingness indicators are artifacts of passive collection

Missingness indicators are predictive, but lossy representations of sequential data



Longitudinal methods can be lossless,

and therefore potentially more predictive,

but are unwieldly to work with.

Our solution: tools to simultaneously visualize and represent:
https://www.andrew.cmu.edu/user/jweiss2/viz.html

Part II: the inattention to low-risk individuals

- Problem: overlooking those at low risk
- Default approach and problems
- Our solution
  - w.r.t. risk prediction
  - w.r.t. risk factors

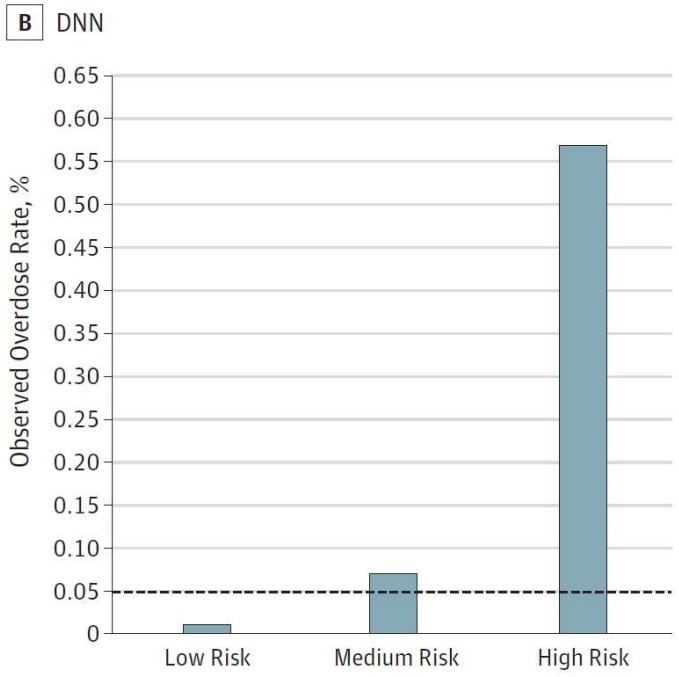
- Case:
  Changes in obtundation in intracerebral hemorrhage patients

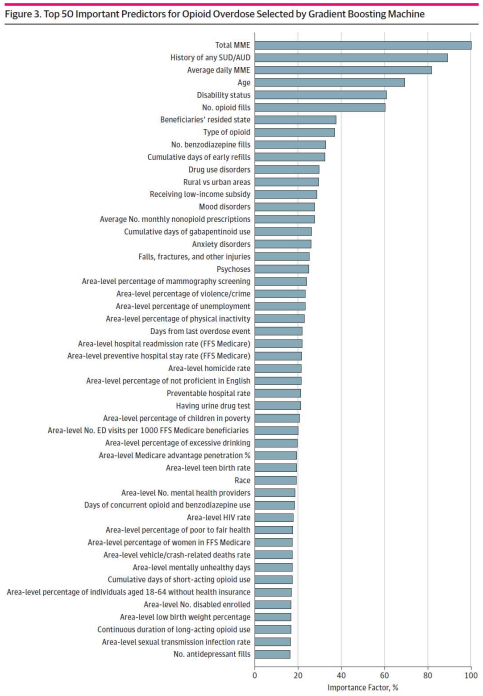
Risk stratification, example 1: opioid overdose

Original Investigation | Substance Use and Addiction

Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions

JAMA Network Open. 2019;2(3):e190968. doi:10.1001/jamanetworkopen.2019.0968





Risk stratification, example 2: sepsis

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis

Table 2. Characteristics of the 4 Phenotypes (continued)

Characteristic[a]	Total	Phenotype			
		α	β	γ	δ
Outcomes					
Mechanical ventilation, median (IQR), d[d]	5 (2-10)	4 (2-9)	4 (2-9)	6 (3-13)	4 (2-9)
Administration of a vasopressor, median (IQR), d[d]	3 (2-5)	2 (2-4)	3 (2-4)	3 (2-5)	3 (2-5)
Admitted to intensive care unit, No. (%)[d]	9063 (45)	1644 (25)	1778 (32)	3381 (63)	2260 (85)
In-hospital mortality, No. (%)	2082 (10)	126 (2)	286 (5)	818 (15)	852 (32)

Risk stratification in the general population

Risk ratios are likely even larger in the general population

e.g. competing risk analyses omit individuals by setting rate = 0

Lower risk individuals are not well characterized

E.g. risk factors for heart attack: based on those at high risk

at-risk women (albeit lower than men) commonly present w/ "atypical" symptoms

Naïve approach: subgroup exploration

Risk factors from linear models for filtration then subsequent optimization make subgroup characterization hard to interpret

How do we build models that produce risk that count individuals equally?

Thought experiment

Two individuals
(1) Risk: 0.2 per year
(2) Risk: 0.02 per year
Absolute change in risk: 0.02
(1) 10% proportional change
(2) 100% proportional change

Survival log likelihood difference:
Near 0, i.e. indifferent to error
on each individual

Implication:
Likelihood-based optimization attends to high risk individuals

When model fails to model risk perfectly (due to limited sample size), it models high risk better, and trains on noise on high risk individuals rather than signal on low risk.

Our approach

Examine the objective function: survival likelihood

Develop a principled alternative

Demonstrate the alternative

Point process likelihood(λ)

Point processes: event rate modeling with rate function λ;
Data: patient := tuples of [**t** time, **x** (event, value)]

Define pdf f, conditioned on events up to t_{i-1}.
For one trajectory (patient) with k target events in time [0, t_k]:

$$L = \prod_{i=1}^{k} f(t_i) = \prod_{i=1}^{k} \lambda(t_i) \exp\Big(- \int_{t_{i-1}}^{t_i} \lambda(u)du \Big) = \Big[\prod_{i=1}^{k} \lambda(t_i) \Big] \exp\Big(- \int_{0}^{t_k} \lambda(u)du \Big)$$

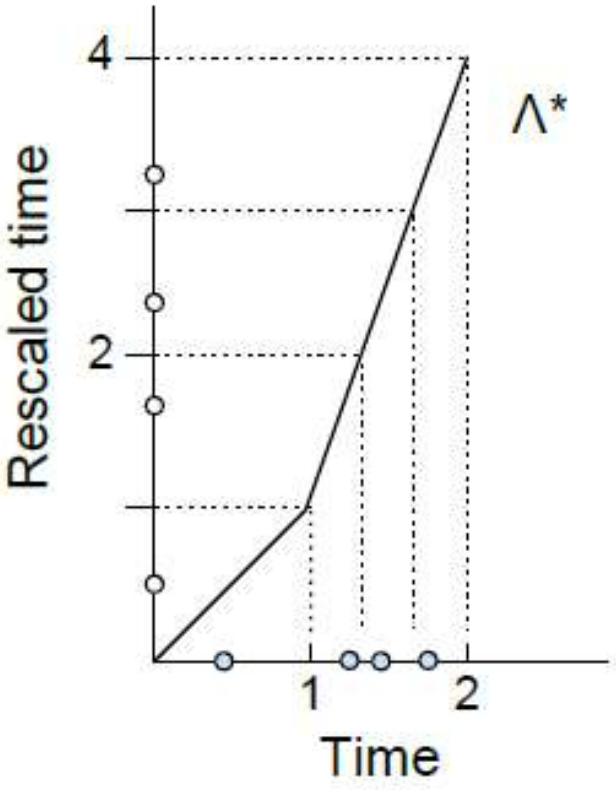
survival term

This quantity is more sensitive to high rates! (large k w.r.t. t)

Recall: $\lambda(t) = \dfrac{f(t)}{1 - F(t)} = \dfrac{f(t)}{S(t)}$

Survival term comes from integrating: $\dfrac{\partial \log S}{\partial t} = \dfrac{\partial S / \partial t}{S(t)} = \dfrac{-f(t)}{S(t)} = -\lambda(t)$

Time rescaling theorem (Meyer 1971, Ogata 1981)



Given Λ* the CDF of λ*, events distributed according to λ* will be distributed according to Poisson(1) in rescaled time

Example (left):

Time 0-1: rate = 1
Time 1-2: rate = 3

equivalent to

Rescaled time 0-4: rate = 1

Implication: attention to high risk

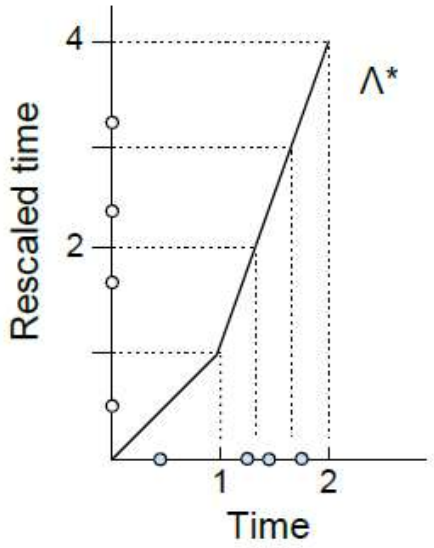
Method

Log likelihood (LL)

$$\mathrm{LL}(\mathbf{X}|\Theta) = \sum_{n=1}^{N}\left(\sum_{i=1}^{T_n}\log\lambda_n(t_{in}|\cdot) - \int_0^{\tau_n}\lambda(t|\cdot)dt\right)$$

Proposed method: adjusted log likelihood (ALL)

$$ALL(\mathbf{X}|\Theta) = \sum_{n=1}^{N}\left(\sum_{i=1}^{T_n}\frac{\log\lambda(t_{in}|\Theta)}{\lambda^*(t_{in})} - \int_0^{\tau_n}\frac{\lambda(t|\Theta)}{\lambda^*(t)}dt\right)$$



We don't have access to the oracle λ*, so we plug-in our predictor, or a lower-variance predictor (avoid dividing by numbers close to 0).

Algorithm 1: Harmonic Mean Point Processes

Result: ALL-trained model

Temporal network $F : X \mapsto [0, \infty)$;

Attention coefficient γ, stability factor ϵ ;

while *training* **do**

  $\hat{\lambda}(t'_{j,k-1}) = F(X_j)$ piecewise on $[t'_{j,k-1}, t'_{j,k}) \; \forall k \in K$;

  Copy then detach $\hat{\lambda}(t'_{j,k-1}) \; \forall j, k$;

  $\mathrm{ALL}_{j,k} = \mathrm{LL}_{j,k} / \left(\hat{\lambda}(t'_{j,k-1})^{\log_{10} \gamma} + \epsilon\right)$;

  ALL.sum().backward();

  optimizer.step();

end

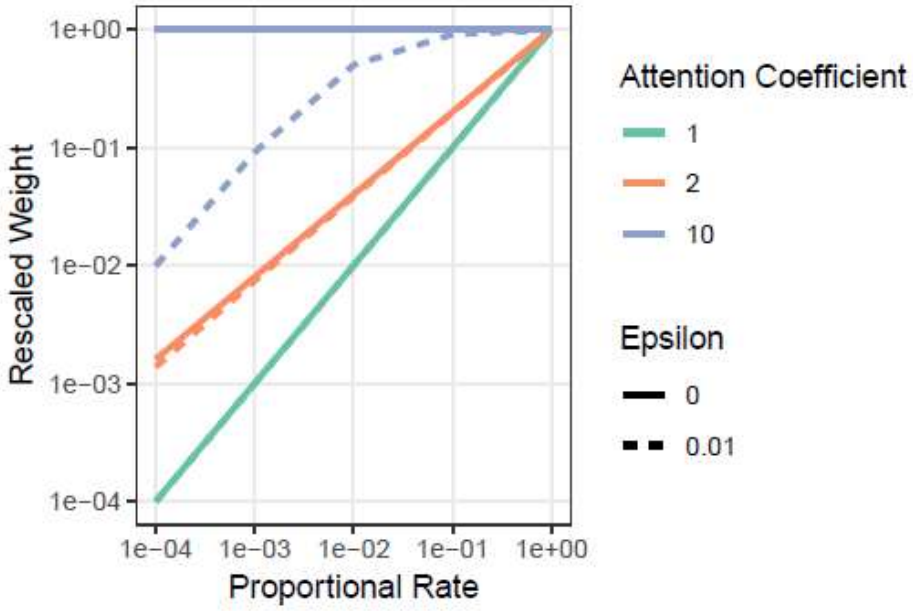
-- while learning a rate model

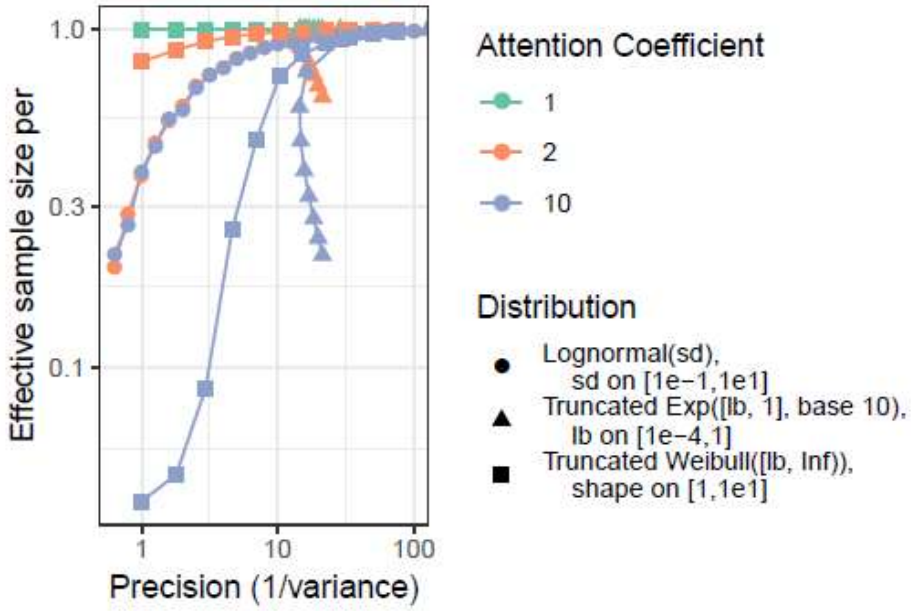
-- compute the rate function

-- reweight by the predicted rate, but don't include the weights in the computation graph

-- update the model

Weight ratio and effective sample size



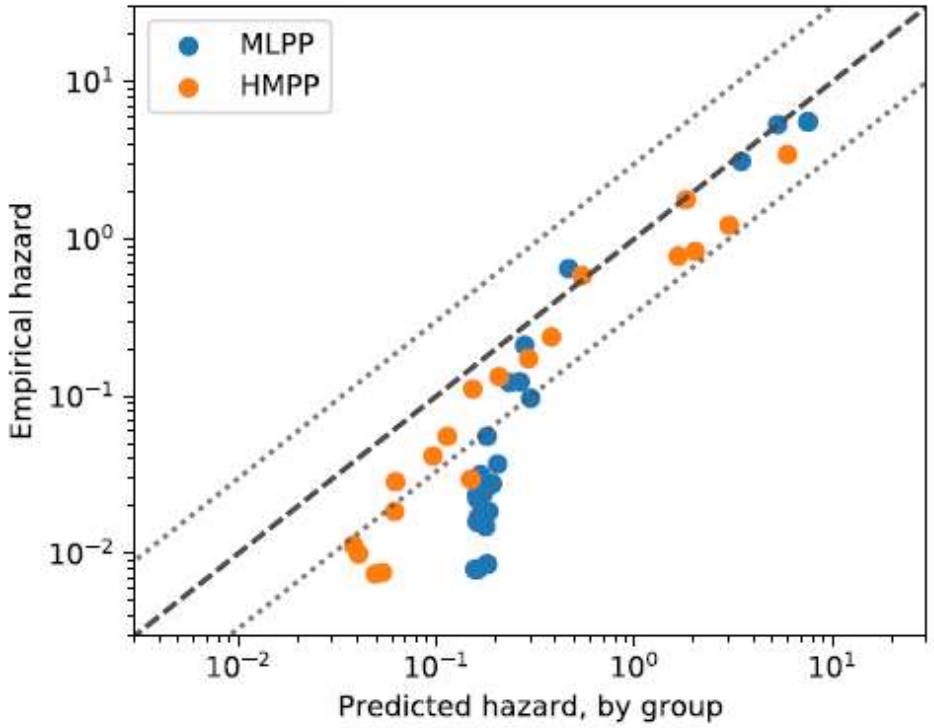


"Without adjustment, low rates are nearly ignored"

"Heavy adjustment leads to very small effective sample sizes"

Simulation

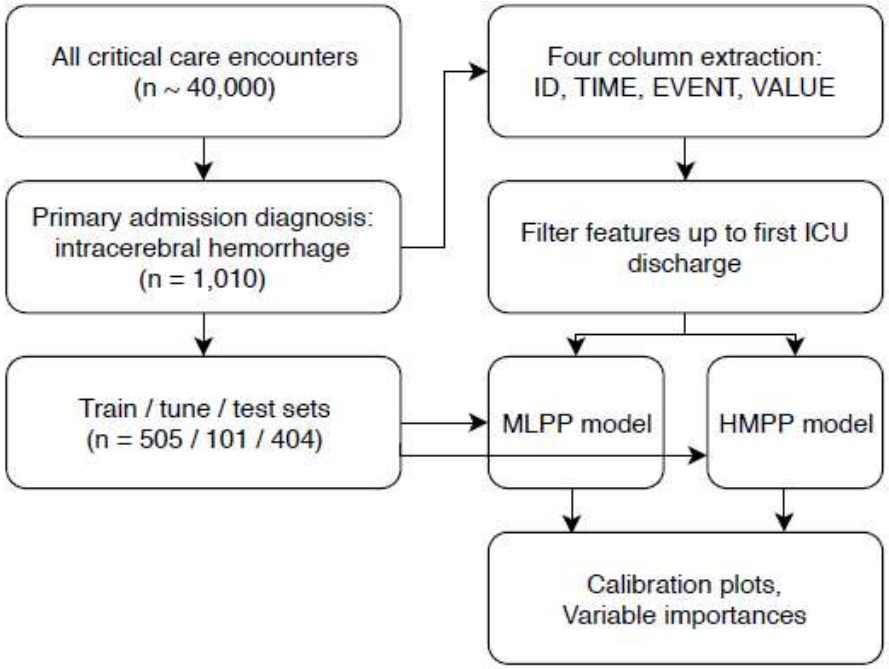
Sample rates λ^* from u ~ Uniform[-2, 1], $\lambda^*=10^u$
Sample trajectories for 10 time units,
with features: (λ^*, t_0), (NA, t_i) otherwise
Train size: 10,000
Test size: 10,000
Model: LSTM



MIMIC III - Prediction of decreasing GCS among ICH patients

ICH: Intracerebral hemorrhage
GCS: Glasgow coma scale (3-15), measure of mental status

MIMIC-III Clinical Database
Alistair Johnson, Tom Pollard, Roger Mark
Published: Sept. 4, 2016. Version: 1.4

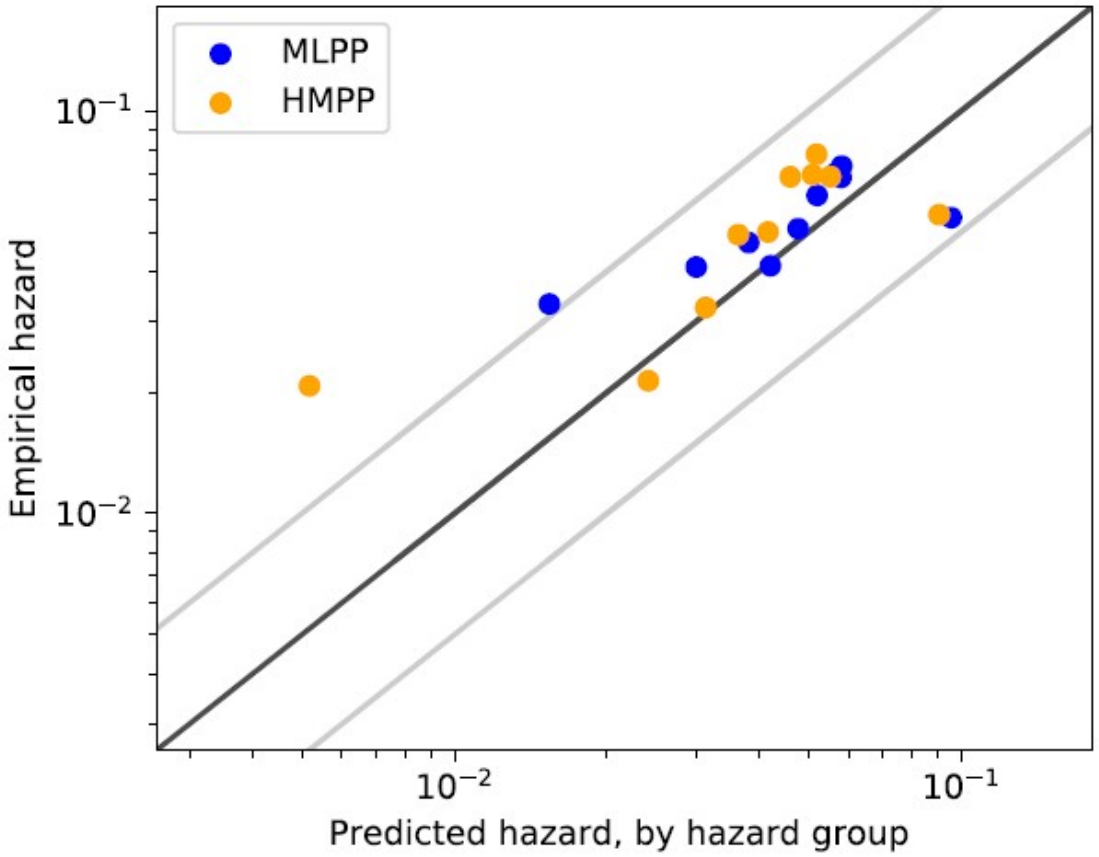


event types = 3810

Feature	*n*=1,010
Age	72 [59, 81]
Gender	
Female	466 (0.46)
Male	544 (0.54)
Ventricular shunt	152 (0.15)
GCS	10 [7, 14]
Decreased GCS, count	3,119
Decreased GCS, rate	1 event / 1.5 days

Use the Wavelet Reconstruction Network architecture from [Weiss 2018]

MIMIC - Prediction of decreasing GCS - Results

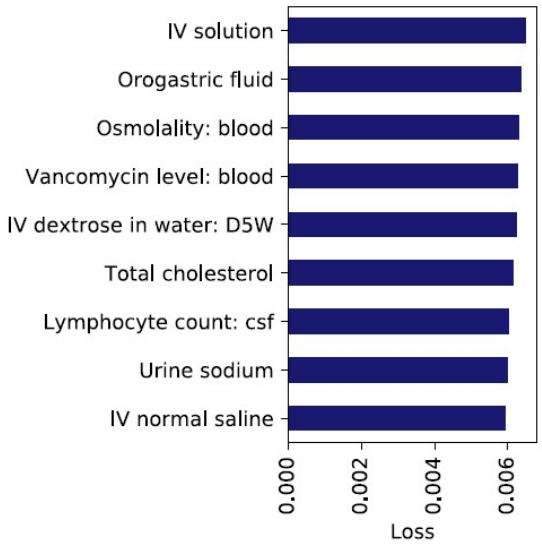


HMPP better isolates low risk subgroups

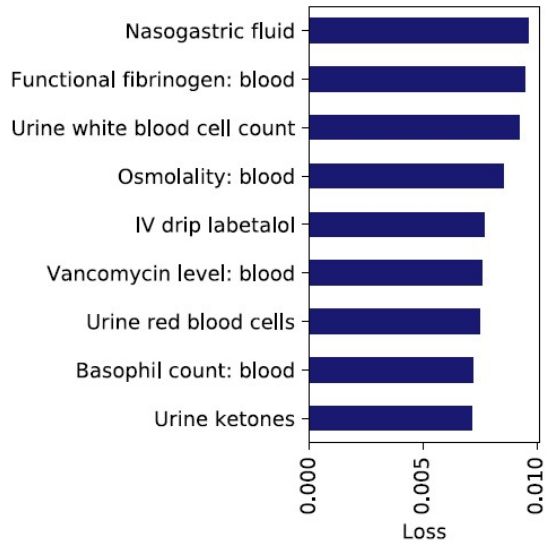
Despite apparent miscalibration, the absolute error is small (0.02 – 0.005 = 0.015)

Prediction of decreasing GCS: variable importance

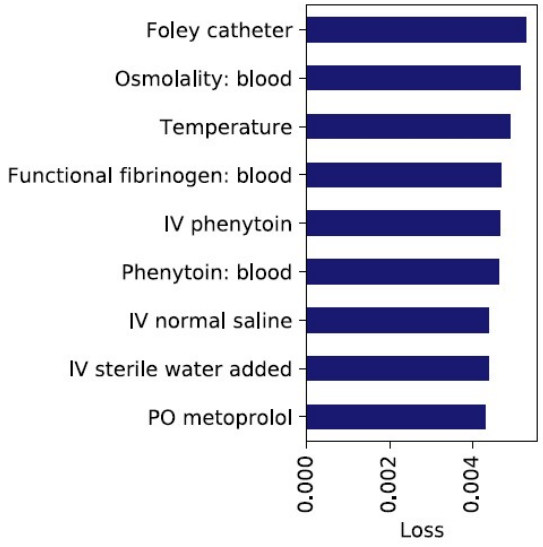
Importance via regularization (L1) loss
Variables deemed important vary by method



Low risk (HMPP)



Max. likelihood (MLPP)



Max. likelihood using HMPP hyperparameters

Conclusions: ML on real-world data (RWD)

RWD is not clean and possibly ill-measured, yet it is more representative

The game is to extract the variation useful for your task
while focusing on use cases where the limitations are not prohibitive

RWD → prediction → risk stratification
RWD→ accept missingness → prediction
RWD → reject missingness → +/- broader populations

RWD and fixed length → lossy representation → limited performance
RWD and fixed length → lossy representation → limited use cases

RWD and fixed length → reject missingness/MAR imputation → broader populations
Up against fine-tuning models site-to-site

Thank you!

Acknowledgments:

Cheng Cheng, Aishwarya Jadhav, Xuejian Wang, Yoonjung Kim
Peggy Peissig, Nathanael Fillmore, Ansh Mehta, George Chen
Chris Seymour, Jason Kennedy, Derek Angus
Weihsuan Jenny Lo-Ciganic, Courtney Kuza, Walid Gellad
Pradeep Ravikumar, Juyong Kim, Linyuan Gong, Justin Khim

CMU / PHDA Center for Machine Learning and Health Awards
CMU - Portugal Award: IntelligentCare
AWS: Amazon Web Services Machine Learning Research Award

Questions and comments:

jeremyweiss@cmu.edu

https://www.andrew.cmu.edu/user/jweiss2/viz.html