Understanding Ventilation from Multivariate ICU Time Series

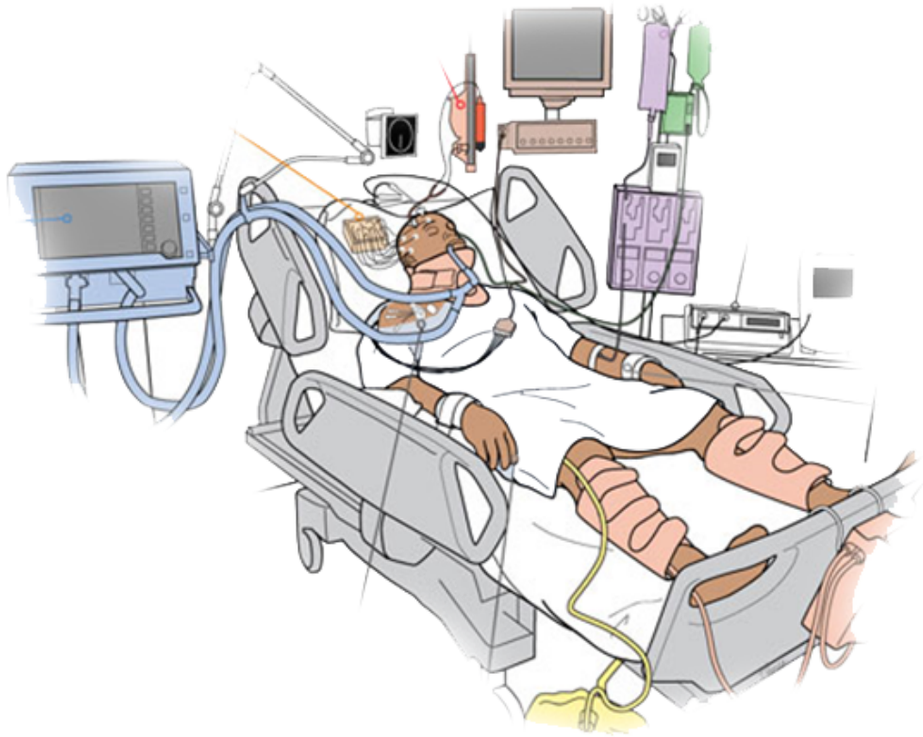
Marzyeh Ghassemi
PhD Candidate
Harvard CMBI Trainee
CSAIL MIT

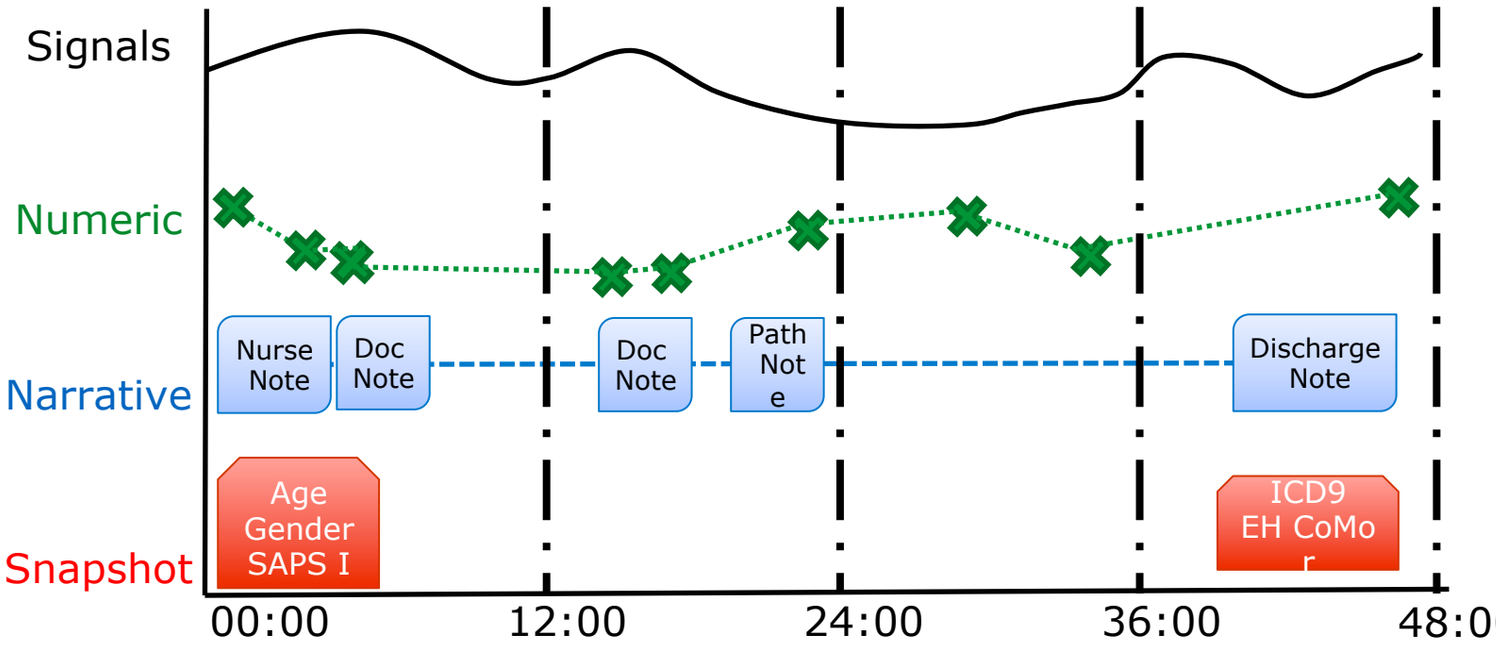
Marco A.F. Pimentel, Mengling Feng,
Finale Doshi, Leo Celi, Peter Szolovits

Massachusetts Institute of Technology

We've Got A Really Big Problem

- ICUs are busy
- Carestaff are inundated with information
- Which patient needs what care?





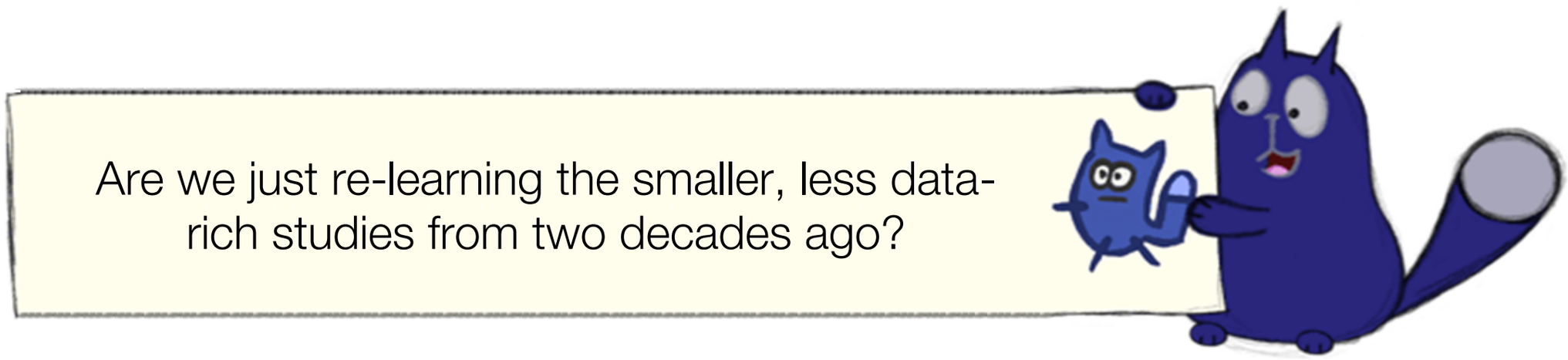
Idea #1: Predict Hospital Mortality?

Pro

- Lots of people do it
- In 2009, 118 validated mortality prediction tools published[1]

Con

- Not accurate across clinical settings[2]
- Models are retrospective, not “realtime”
- Associations are not really actionable



[1] Siontis, George CM, Ioanna Tzoulaki, and John PA Ioannidis. "Predicting death: an empirical evaluation of predictive tools for mortality." *Archives of internal medicine* 171.19 (2011): 1721-1726.
[2] Grady, Deborah, and Seth A. Berkowitz. "Why is a good clinical prediction rule so hard to find?." *Archives of internal medicine* 171.19 (2011): 1701-1702.

Idea #2: Predict Interventions?

Pro

- The ICU is playing an expanding role in acute hospital care[1]
- The value of many treatments and interventions in the ICU is unproven[2], and high-quality data supporting or discouraging specific practices are sparse
- Many standard treatments are ineffective, or even harmful to patients[3]

Con

- Treatments are constantly being given/changed
- Effects of different interventions are not isolated
- Difficult to separate the *intent* to intervene from the *need* to intervene

[1] Vincent, Jean-Louis. "Critical care-where have we been and where are we going." Crit Care 17.Suppl 1 (2013): S2.
[2] Vincent, Jean-Louis, and Mervyn Singer. "Critical care: advances and future perspectives." *The Lancet* 376.9749 (2010): 1354-1361.
[3] Ospina-Tascón, Gustavo A., Gustavo Luiz Büchele, and Jean-Louis Vincent. "Multicenter, randomized, controlled trials evaluating mortality in intensive care: Doomed to fail?." *Critical care medicine* 36.4 (2008): 1311-1322.

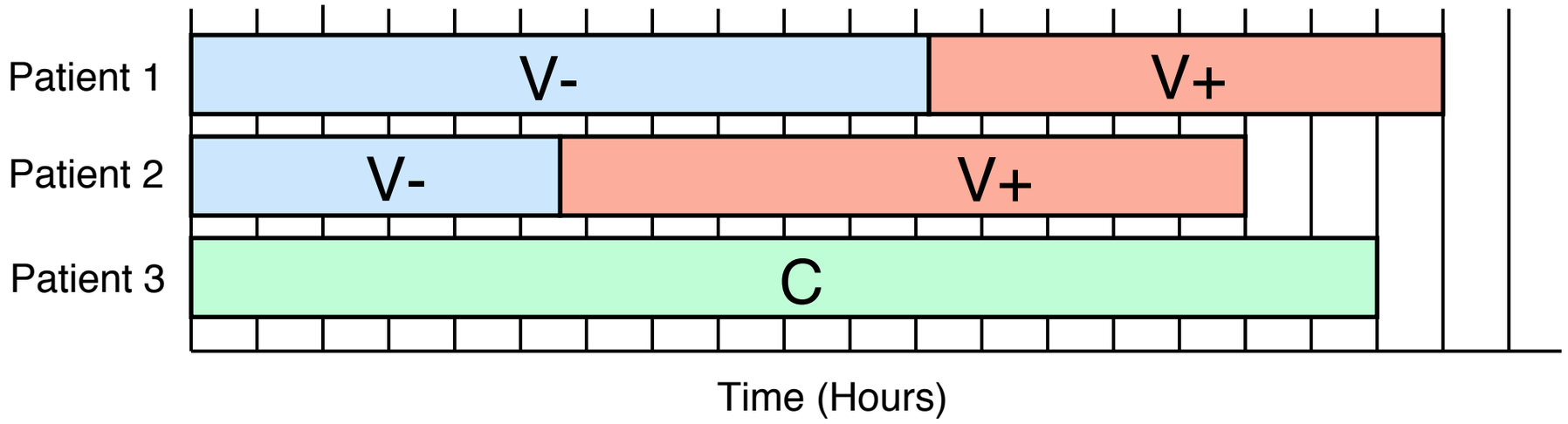
Idea #3: Understanding Intervention Effects

- Use data to understand effects of the ICU practices
- Not possible with conventional observational studies (regardless of size)
- Many possible interventions, focus on ventilation
  - Clinicians are continually trying to predict the earliest time that a patient can resume spontaneous breathing[1]
  - Small changes in the timing and setting of the ventilation can make large differences in patient outcomes[2]

[1] Yang, Karl L., and Martin J. Tobin. "A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation." New England Journal of Medicine 324.21 (1991): 1445-1450.
[2] Tobin, Martin J. "Principles and practice of mechanical ventilation." (2006): 426.

Study Goals

- Explore differences in multidimensional physiological timeseries for ICU populations:
  - pre-ventilation (V-)
  - post-ventilation (V+) and
  - non-ventilated (C)



- If we can recover ventilation state, could lead to a better understanding of the need vs. “intent” to intervene.

Study Data

- Patients with 12:96 hours of data in MIMIC2v26, always full code - 6,855 patients.

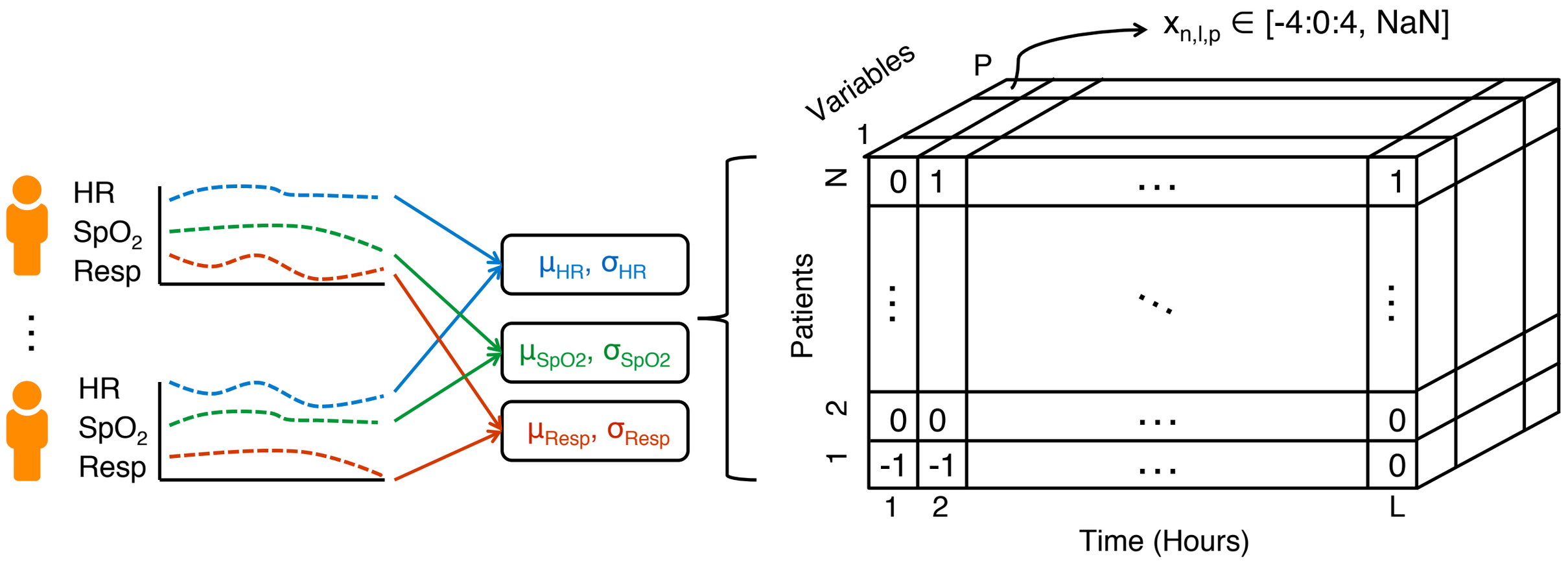
- Numeric physiological time series data where >10% of patients had >40% recording instances.
  - hematocrit (HCT)
  - heart-rate (HR)
  - mean arterial blood pressure (MeanBP)
  - blood oxygenation level (SPO2)
  - temperature (TEMP)
  - spontaneous respiration rate (RESP).
  - bicarbonate (BICAR)
  - potassium (K)
  - sodium* (Na)
  - glucose* (GLU)

- Any modification of ventilation settings is an indicator of ventilation in the hour it occurred. Ventilation gaps < 6 hours are continuous.

	Ventilated V-/V+	Control (C)
Patients	3401	3454
Hours	14116 V- 213017 V+	225217
Age	63.16 ± 16.41	62.46 ± 17.62
% Male	60 %	54%
Hospital LoS	9.53 ± 8.02	9.46 ± 8.99
Hospital Mort.	4%	3%
ICU LoS	2.83 ± 0.63	3.48 ± 14.56
ICU Mort.	3%	1%
SAPS-I	16.47 ± 4.26	10.55 ± 3.81
Admit SOFA	7.96 ± 2.98	3.00 ± 2.37
MICU	23.14%	49.02%
SICU	23.38%	24.75%
CCU	8.62%	21.71%
CSRU	44.87%	4.52%

Data Preprocessing

- Collect timeseries over *N* patients, *L* hourly timesteps and *P* variables
- Z-score and discretize timeseries using population μ, σ.



- Every $x_{n,l,p}$ is one of ten possible characters, -4:0:4 or NaN.
- Every $x_{n,l}$ is one of 10^P possible words.

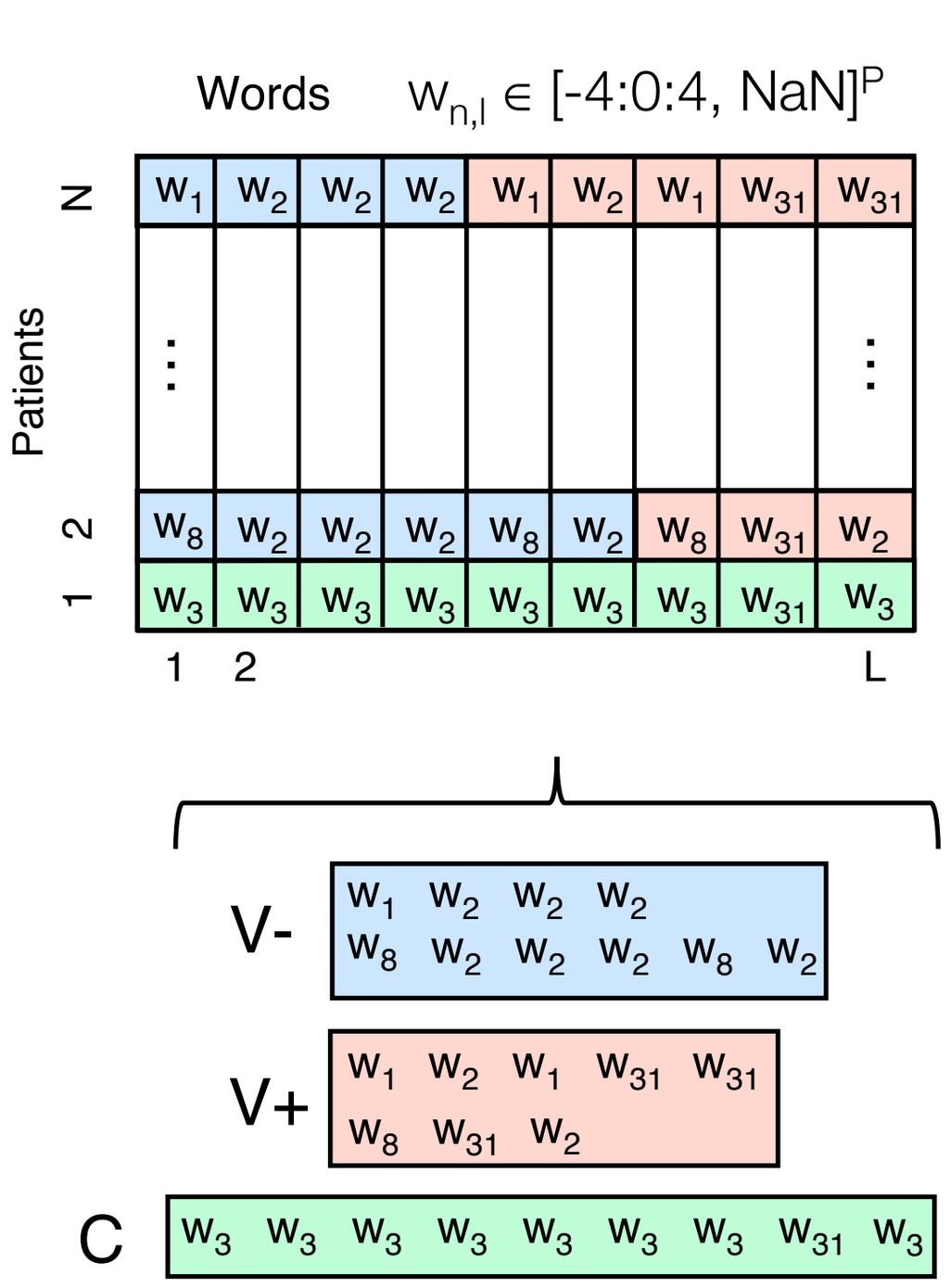
Statistical Language Model

- If each hourly vector is a “word”, where words ∈ [-4:0:4, NaN]^P e.g. for P = 4, w_1 = [0 0 0 1]

- Our 12-96 hour x 6,855 patient matrix ~ a set of 6,855 sentences.

- Mapping this onto a LM framework, we estimate $P(w_i \mid w_{i-1}, \ldots, w_{i-t})$ for some number of t grams.

- Representation is generative, works well for understanding sequences.



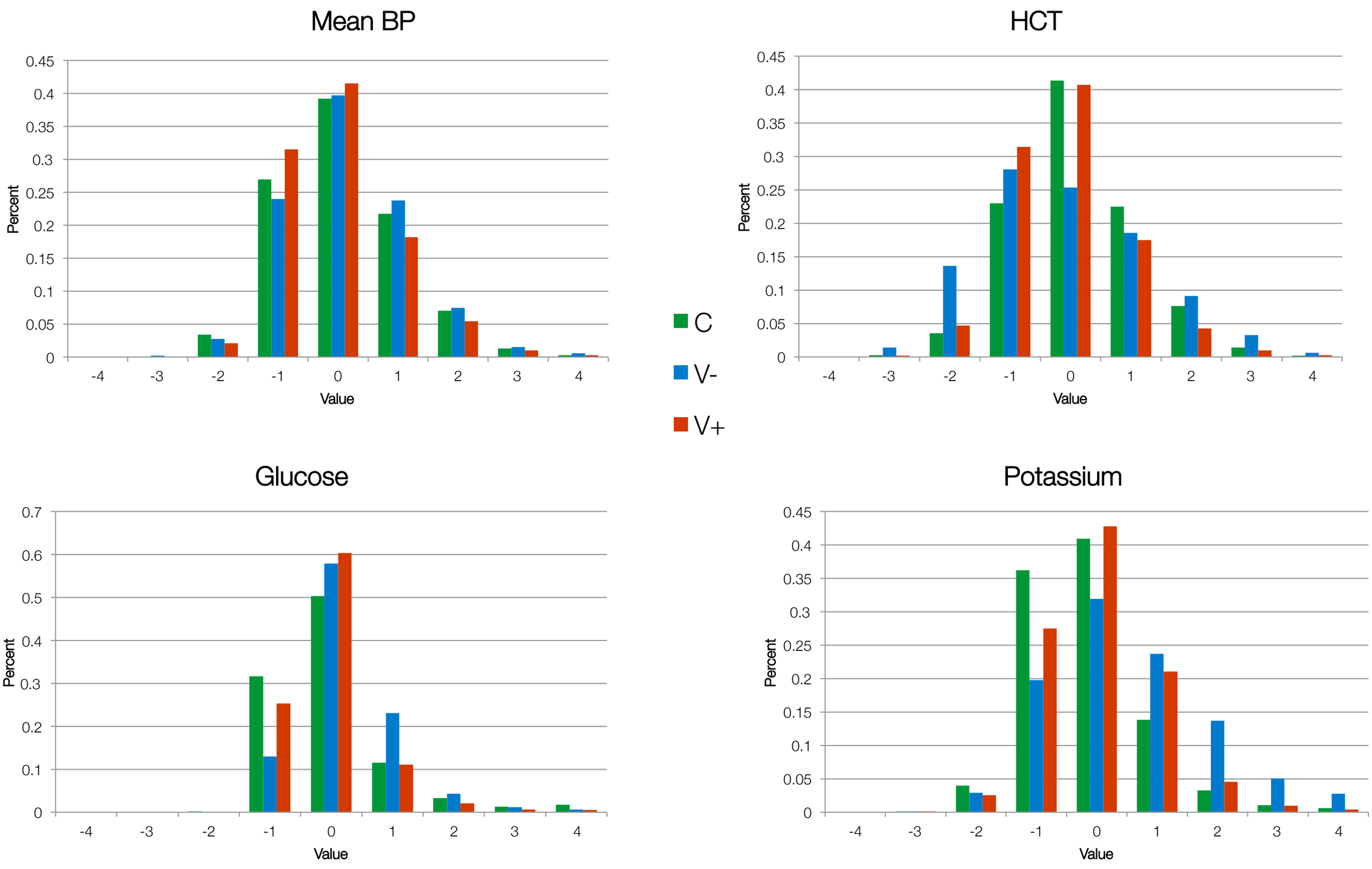
Language Model Training/Evaluation

- Generate per-class (V-/V+/C) text streams, and build per-class LMs
  - Vocabulary is the 20,000 most common words
  - 70/30% train/test splits
  - Trigram LM with modified Kneser-Ney[1] smoothing

- Measure the perplexity of the language models in test set
  - The LM is a probability distribution *q* over sentences
  - On test samples $x_1 \ldots x_T$,
    q has entropy $H = -\sum q(x) \log_2 q(x)$,
    perplexity = 2^H

[1] Chen, Stanley F., and Joshua Goodman. "An empirical study of smoothing techniques for language modeling." *Computer Speech & Language* 13.4 (1999): 359-393.

Aggregate Class Differences



“Languages” Follow Intuitive Trends

hr	meanbp	spo2	temperature f	spontaneous rr	Count
0	0	0	0	0	3904
-1	0	0	0	0	3737
-1	0	0	0	-1	3093
0	0	0	0	-1	2631
-1	-1	0	0	0	2620
0	-1	0	0	0	2589
0	1	0	0	0	2493
0	0	1	0	0	2376
0	0	0	0	1	2152

Control

hr	meanbp	spo2	temperature f	spontaneous rr	Count
-1	0	0	0	0	139
-1	0	0	0	-1	127
-1	0	1	0	-1	117
0	0	0	0	0	102
-1	1	0	0	-1	97
-1	0	1	0	0	80
0	0	0	0	-1	77
0	0	1	0	-1	75
0	1	0	0	0	72

Unvent

hr	meanbp	spo2	temperature f	spontaneous rr	Count
0	0	0	0	0	4000
0	0	1	0	-1	3492
0	-1	0	0	0	3254
0	0	1	0	0	3034
0	0	0	0	-1	2854
0	-1	0	0	-1	2497
0	-1	1	0	-1	2448
0	0	0	0	1	2421
-1	0	1	0	-1	2297

Vent

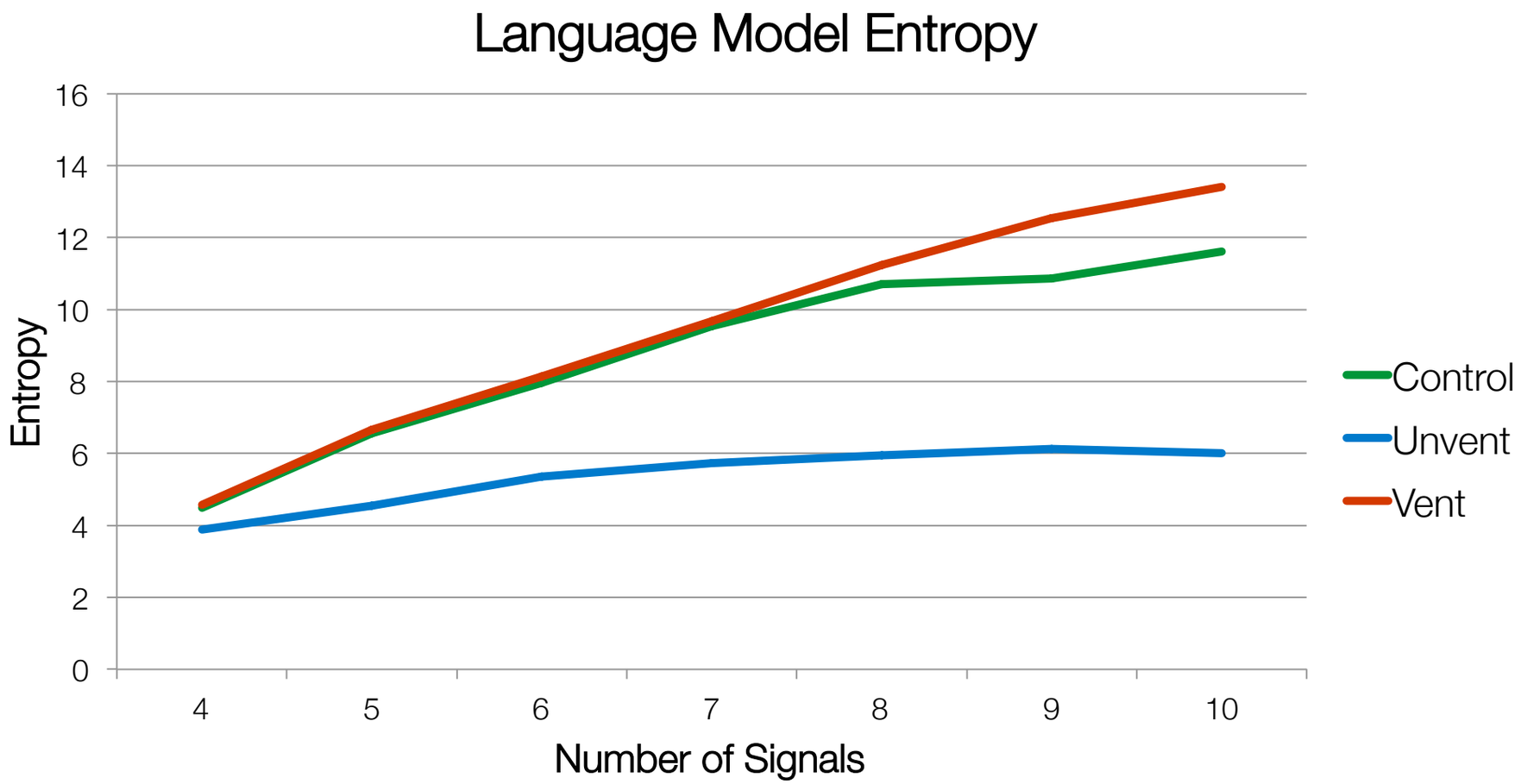
- Word frequencies roughly followed a power-law distribution
- Most common words correspond to physiological stability, or depressed heartrate, blood pressure, and respiration.
- Labs values are more often missing, leading to more sparsity ($6{,}328/10^5$ vs $164{,}182/10^{10}$ unique words observed, or 0.0633 vs. 0.0000164)

Per-class Sequence Prediction

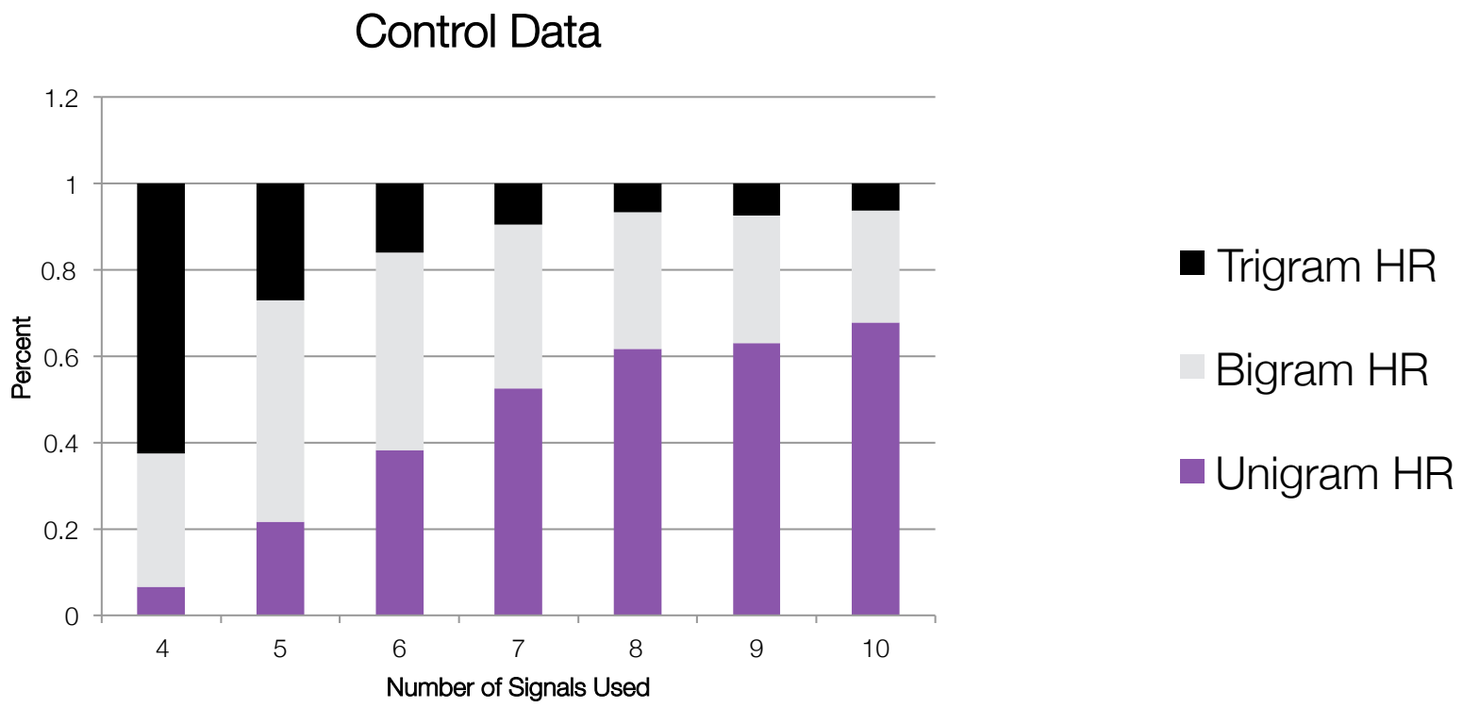
Metric	V-	V+	C
Perplexity	44.12	19.40	18.72
Entropy	5.46	4.28	4.23
3-gram HR	34.66%	68.31%	67.65%
2-gram HR	42.76%	18.29%	18.47%
1-gram HR	22.58%	13.40%	13.88%

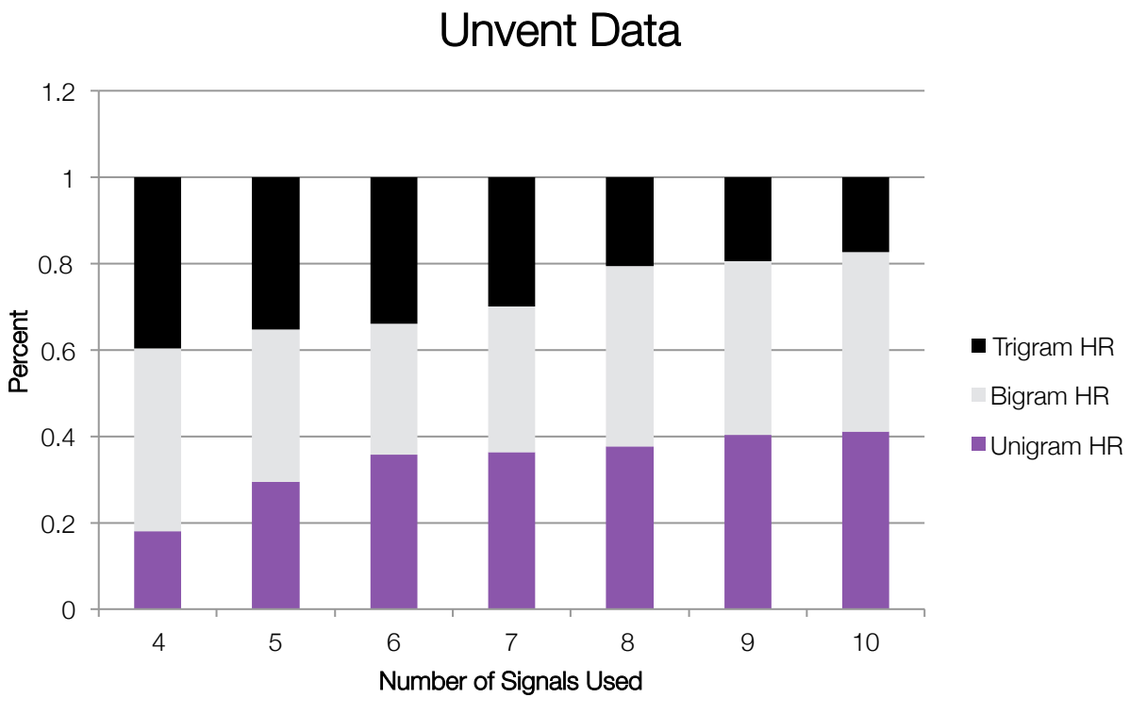
If we use only the four most common signals, it seems like post-ventilated and control patients have similar sequence perplexity.

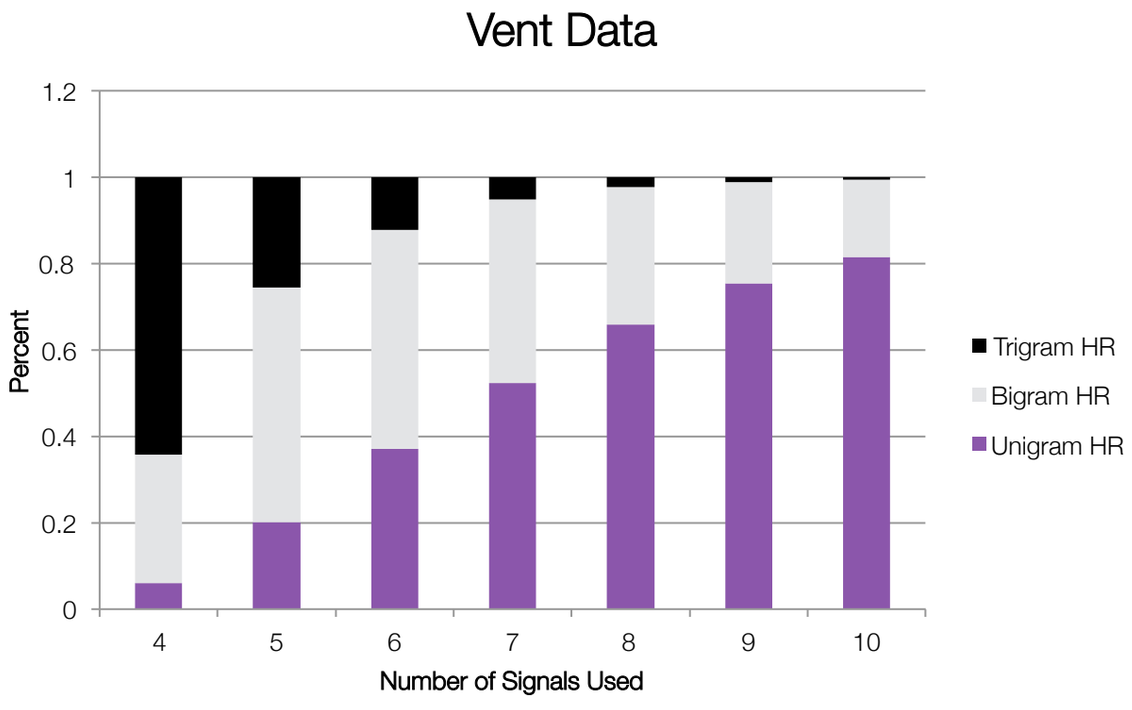
LM Perplexity Changes with More Interpolation



Gram Hit-rate Varies Depending on Data Used





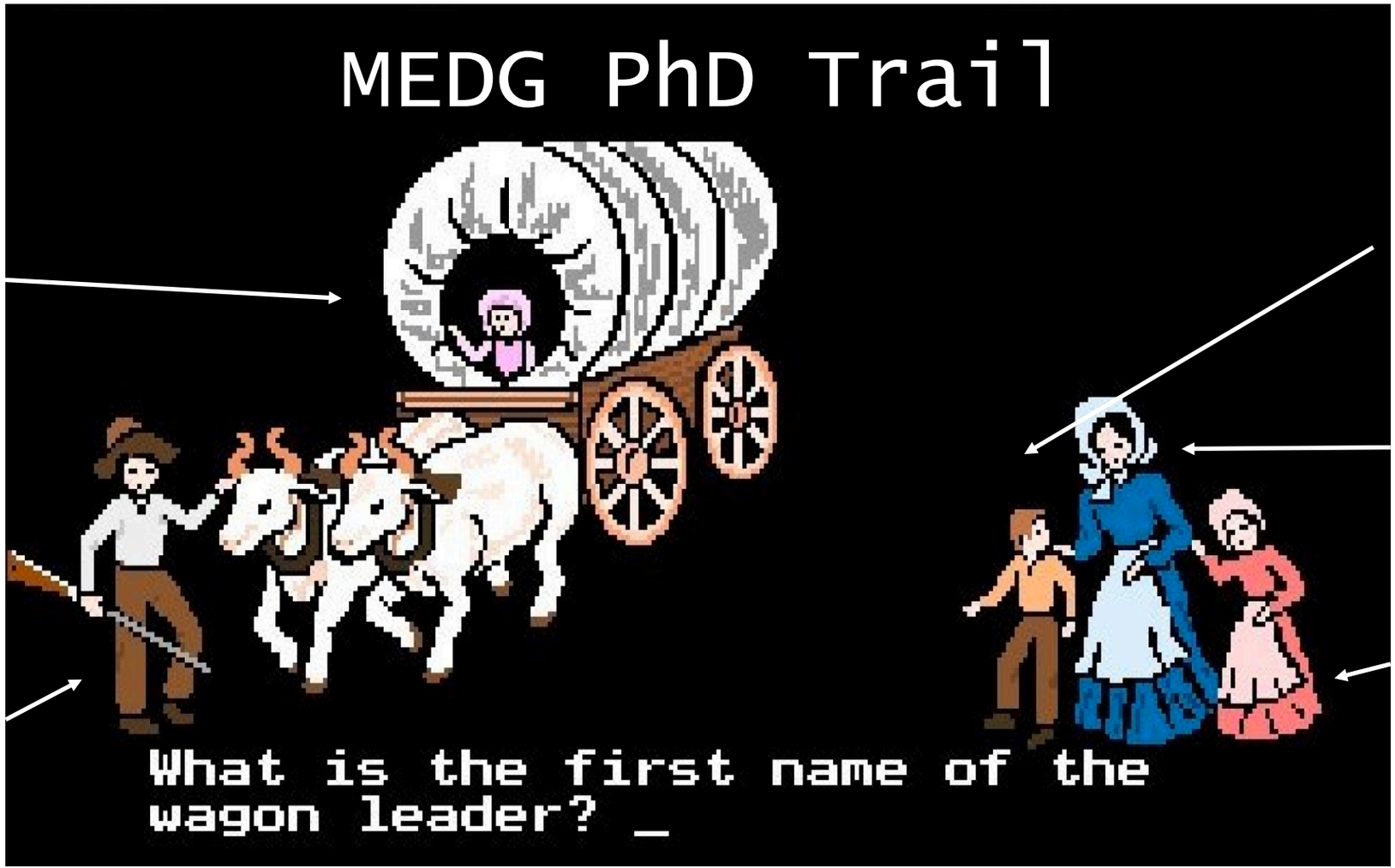


Future Work

- Look at sequence prediction
- Building factored language models
- Examining other ICU interventions
- Examining other non-ICU interventions

Thanks!

Intel Science and Technology Center for Big Data
National Library of Medicine Biomedical Informatics Research Training grant (NIH/NLM 2T15 LM007092-22)
R01 grant EB001659 from the NIBIB of the NIH
RCUK Digital Economy Programme
A*STAR Graduate Scholarship.



Tristan Naumann

Yuan Luo

Marzyeh Ghassemi

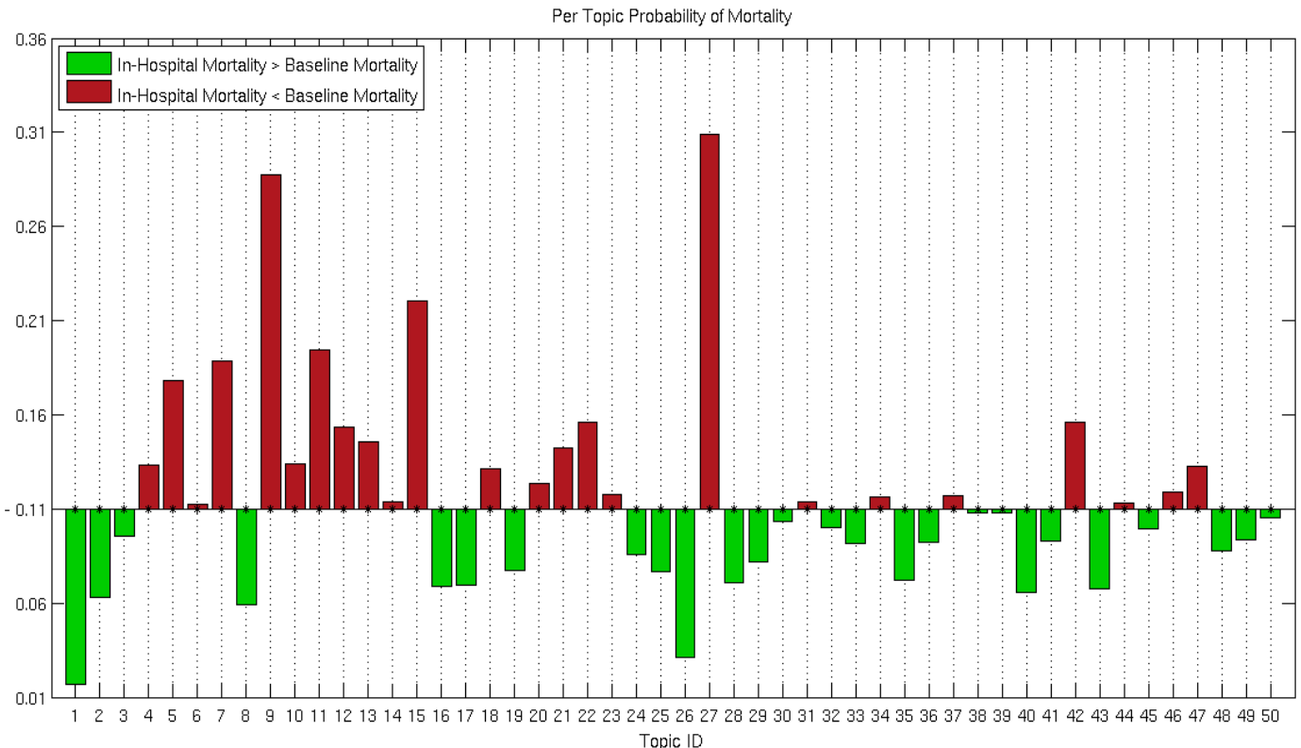
Lydia Letham

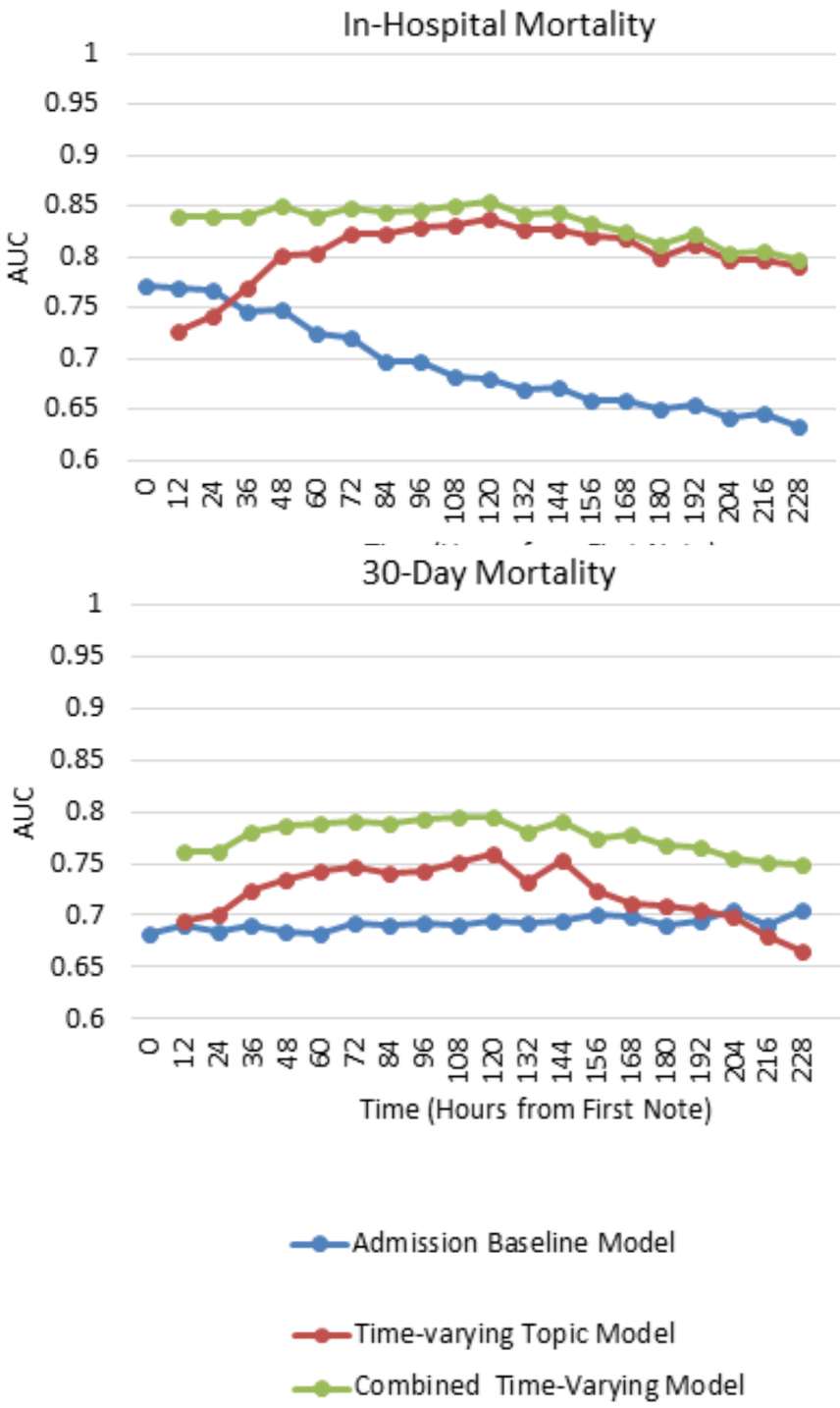
Dr. Peter Szolovits

Backup

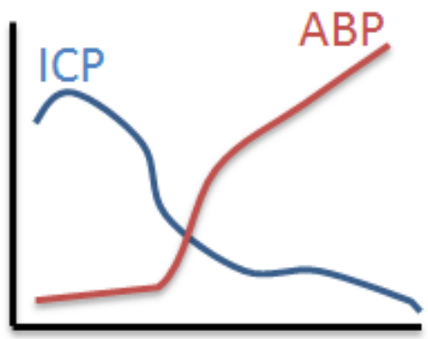
KDD 2014 - Unfolding Physiological State: Mortality Modelling in Intensive Care Units

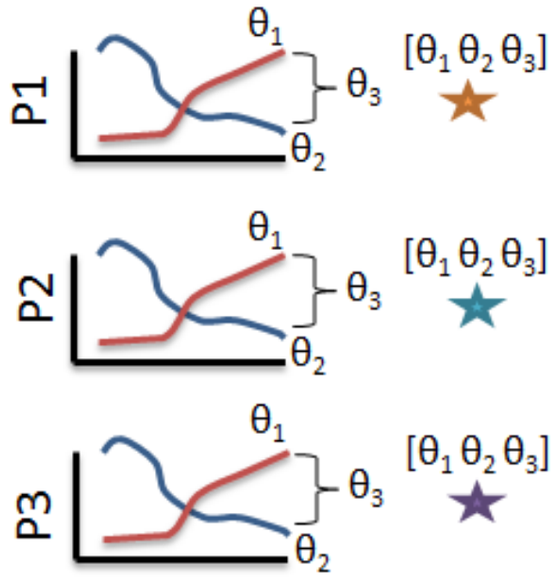
	Topic #	Top Ten Words	Possible Topic
In-Hospital Mortality	27	name family neuro care noted	Discussion of end-of-life care
	15	intubated vent ett secretions propofol	Respiratory failure
	7	thick secretions vent trach resp	Respiratory infection
	5	liver renal hepatic ascites dialysis	Renal failure
Hospital Survival	1	cabg pain ct artery coronary	Cardiovascular Surgery
	40	left fracture ap views reason	Fracture
	16	gtt insulin bs lasix endo	Chronic diabetes
1 Year Mortality	3	picc line name procedure catheter	PICC line insertion
	4	biliary mass duct metastatic bile	Cancer treatment
	45	catheter name procedure contrast wire	Coronary catheterization

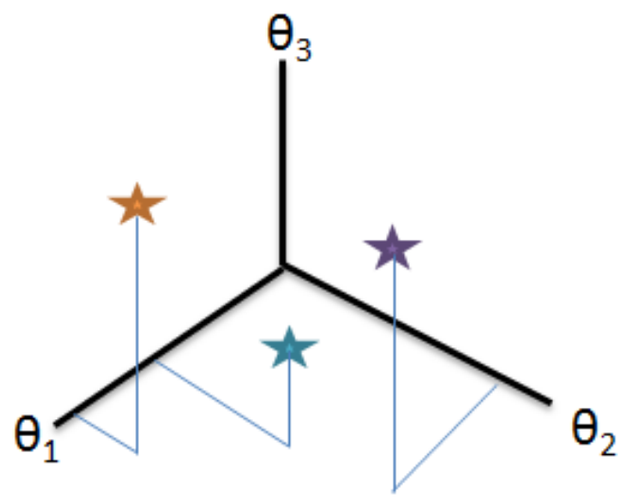




Idea #2: Latent Acuity *through* Hospital Mortality!







Clinical Signals → **MTGP Hyperparameters** → **Latent Space**

MTGP models capture the movement within and between signals.

We transform a variety of irregularly-sampled signals into a new latent space using MTGP hyperparameters.

Patients are compared based on their similarity in the new latent space.

θ provides a **new latent search space** to examine and evaluate the **similarity** of any two given **multi-dimensional functions.**

AAAI 2015 - A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU with Sparse, Heterogeneous Clinical Data

- Case Study #1: Forecast the MAP and ICP signals & estimate cerebrovascular pressure reactivity (PRx) in TBI patients

Signal	Measure	STGP	MTGP
ICP	RMSE	0.91	0.69
	MSLL	0.6	0.45
ABP	RMSE	2.77	1.98
	MSLL	0.65	0.55
PRx-PRx*	RMSE	-	0.09

- Case Study #2: Using MTGP hyperparameters as additional classification features for mortality prediction

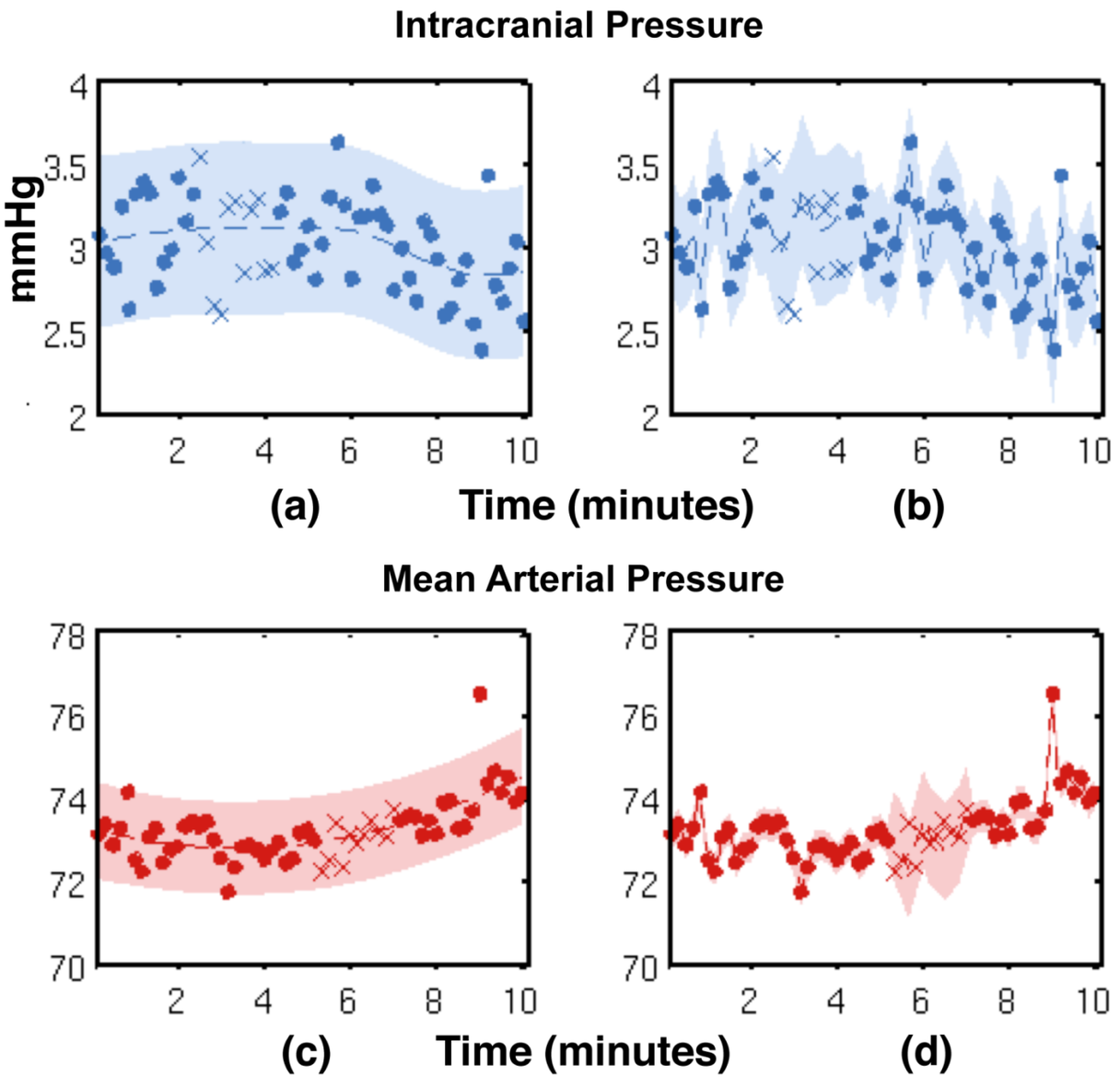
Features	Hospital Mortality	1-Year Mortality
SAPS-I	0.702	0.500
Ave. Topics	0.759	0.653
SAPS-I + MTGP	0.775	0.624
Ave. Topics + MTGP	0.788	0.673
SAPS-I + Ave. Topics + MTGP	0.812	0.686

Case Study #1:
Estimating Signal in Traumatic Brain Injury

- ICP and ABP data collected from 35 TBI patients who were monitored for 24+ hours in a Neuro-ICU.

- Our goal was to forecast the MAP and ICP signals as well as estimate cerebrovascular pressure reactivity (PRx)

Signal	Measure	STGP	MTGP
ICP	RMSE	0.91	0.69
	MSLL	0.6	0.45
ABP	RMSE	2.77	1.98
	MSLL	0.65	0.55
PRx-PRx*	RMSE	-	0.09



Modified Kneser-Ney Model

- Intuition: how likely a word w_i is to appear + how likely it is to appear in an unfamiliar t-gram context.
  - Use interpolation instead of backoff.
  - Use a separate discount for one/two-counts instead of a single discount for all counts.
- Estimates discounts on held-out data instead of using a formula based on training counts.

$$P_{KN}(w_i \mid w_{i-1}) = \frac{\max(c(w_{i-1}w_i) - \delta, 0)}{c(w_{i-1})} + \lambda \frac{|\{w_{i-1} : c(w_{i-1}, w_i) > 0\}|}{|\{w_{j-1} : c(w_{j-1}, w_j) > 0\}|}$$

$$\lambda(w_{i-1}) = \frac{\delta}{c(w_{i-1})} |\{w' : c(w_{i-1}, w') > 0\}|$$