Neural Topic Models with Survival Supervision

Jointly Predicting Time-to-Event Outcomes and Learning How Clinical Features Relate

Linhong Li[1] Ren Zuo[2] Amanda Coston[1] Jeremy C. Weiss[1] George H. Chen[1]

[1] Carnegie Mellon University

[2] Cornerstone Research

International Conference on Artificial Intelligence in Medicine 2020

Motivation

Survival analysis

Goal. Predict time-to-event outcomes (e.g., time until death, length of stay in ICU)

In clinical applications that demand an interpretable survival model, standard approach: use Cox proportional hazards (Cox 1972), possibly with regularization/variable selection

In typical use, does not learn how features relate

(can manually add pairwise interactions but this gets costly for large # of features)

Goal. Discover "clusters" of features that co-occur

- Analogous to how clinicians look at *constellations of symptoms* (called *syndromes*)
- Accommodate large # of features

Topic modeling

This paper: New neural net framework for combining topic modeling and survival analysis that retains interpretability

How Our Paper Relates to Existing Literature

A topic model with survival supervision already exists (Dawson & Kendziorski 2012)

- Combines the latent Dirichlet allocation (LDA) topic model (Blei et al 2003) with the Cox proportional hazards survival model (Cox 1972)
- Learns the joint topic/survival model via variational inference
    - Their algorithm does not scale to large datasets
    - Their algorithm is "bespoke" — changing which topic or survival model is used would require re-deriving significant portions of the inference algorithm

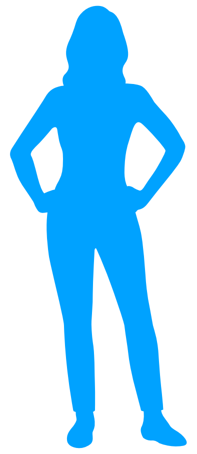
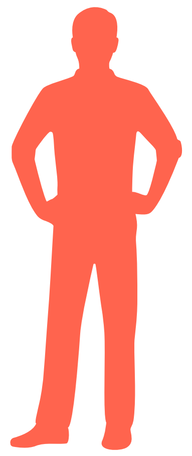
Our new neural network framework combines <u>any</u> topic model and <u>any</u> survival model that have neural network formulations

resolves both issues

- Many topic models have neural net approximations/formulations (Srivastava & Sutton 2017, Card et al 2018, Dieng et al 2019, …)
- Many survival models have neural net formulations (Faraggi & Simon 1995, Katzman et al 2018, Lee et al 2018, …)

Survival Analysis

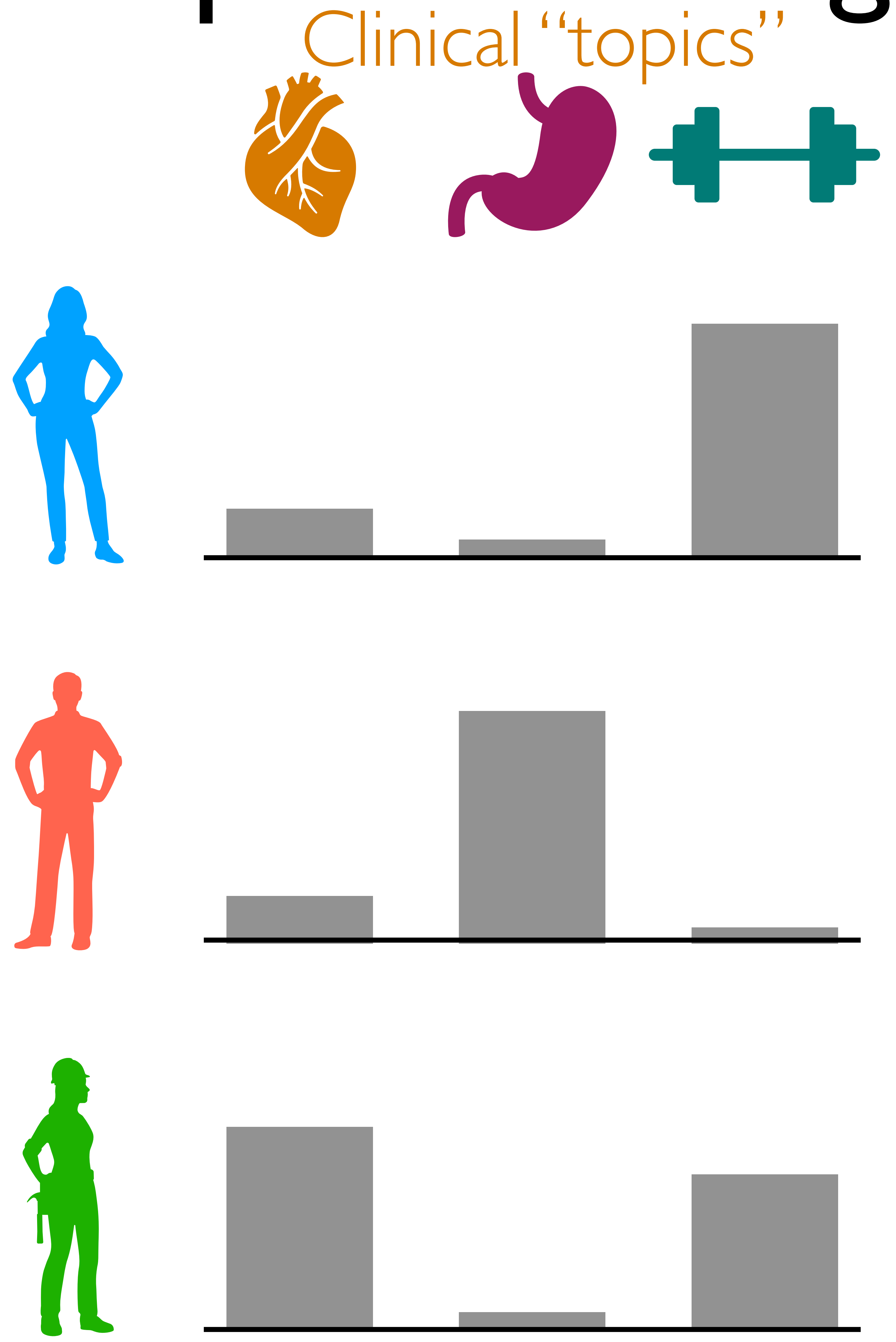
When we stop collecting training data, some subjects are still alive

	Gluten allergy	Immuno-deficiency	Low resting heart rate	Irregular heart beat	High BMI	Time until death
	✗	✗	✓	✗	✗	2 days
	✓	✓	✗	✗	✓	10 days
	✗	✗	✗	✓	✗	≥ 6 days

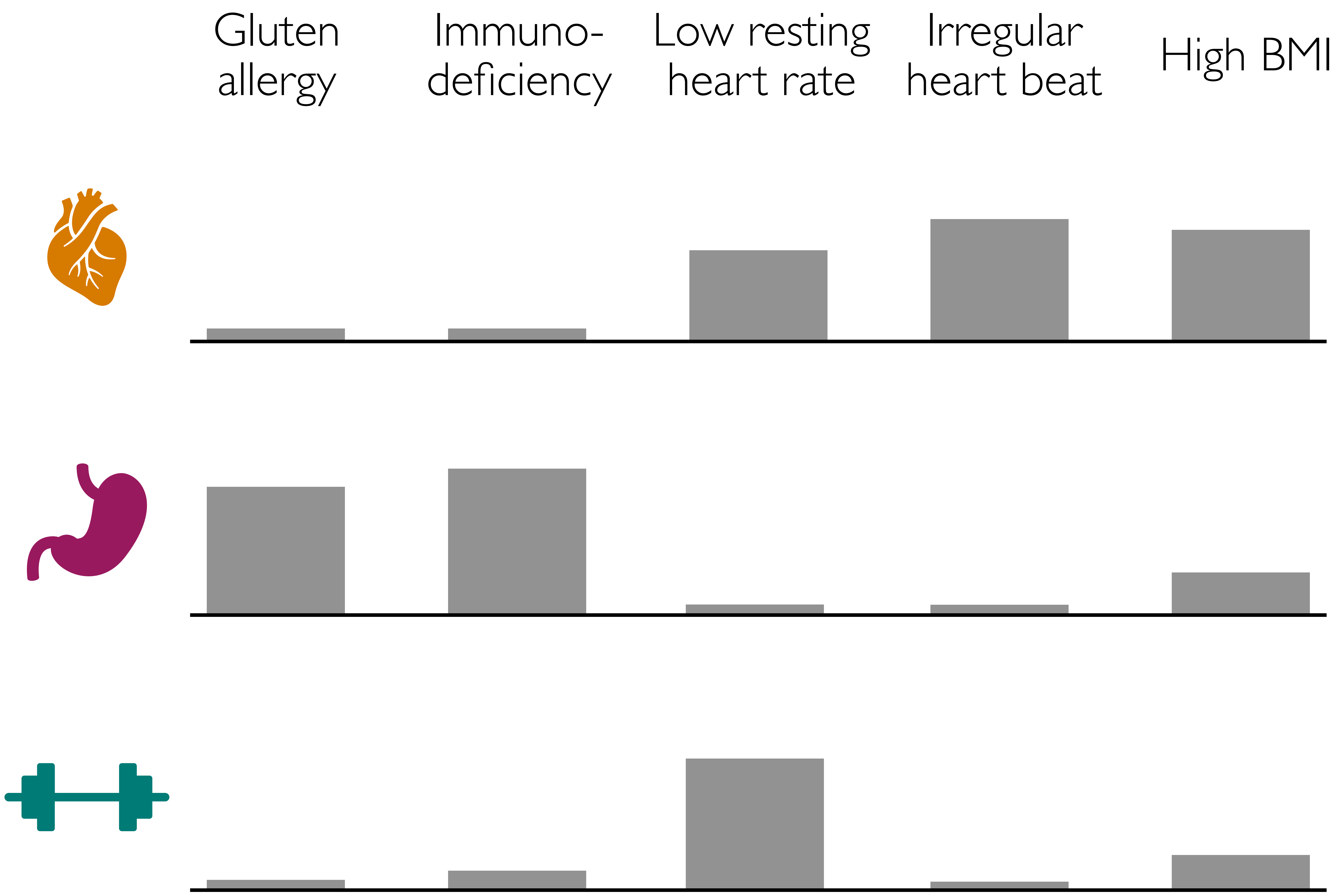
Feature vectors (for simplicity, each feature is binary)

Time-to-event outcome to predict

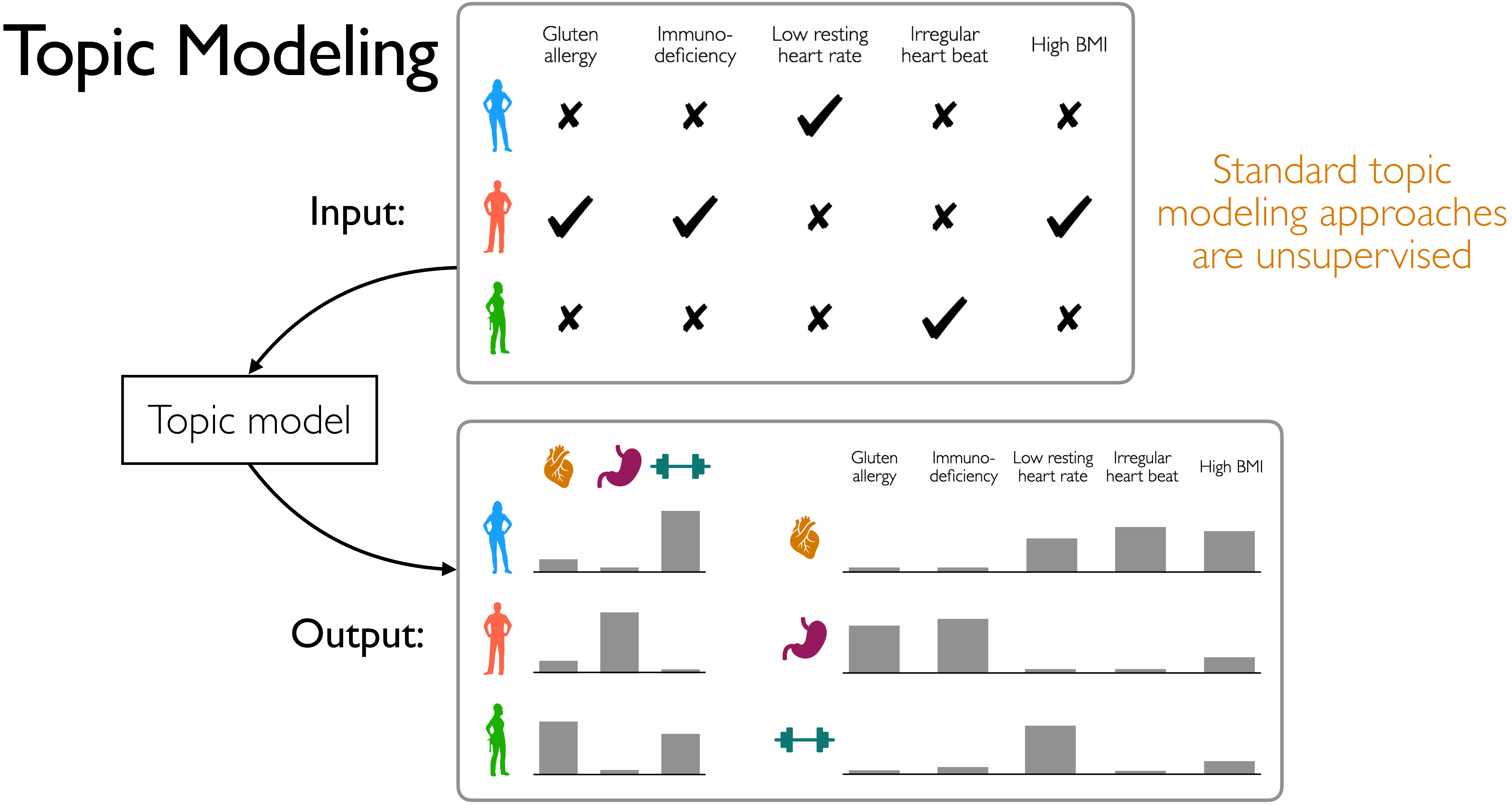
Topic Modeling

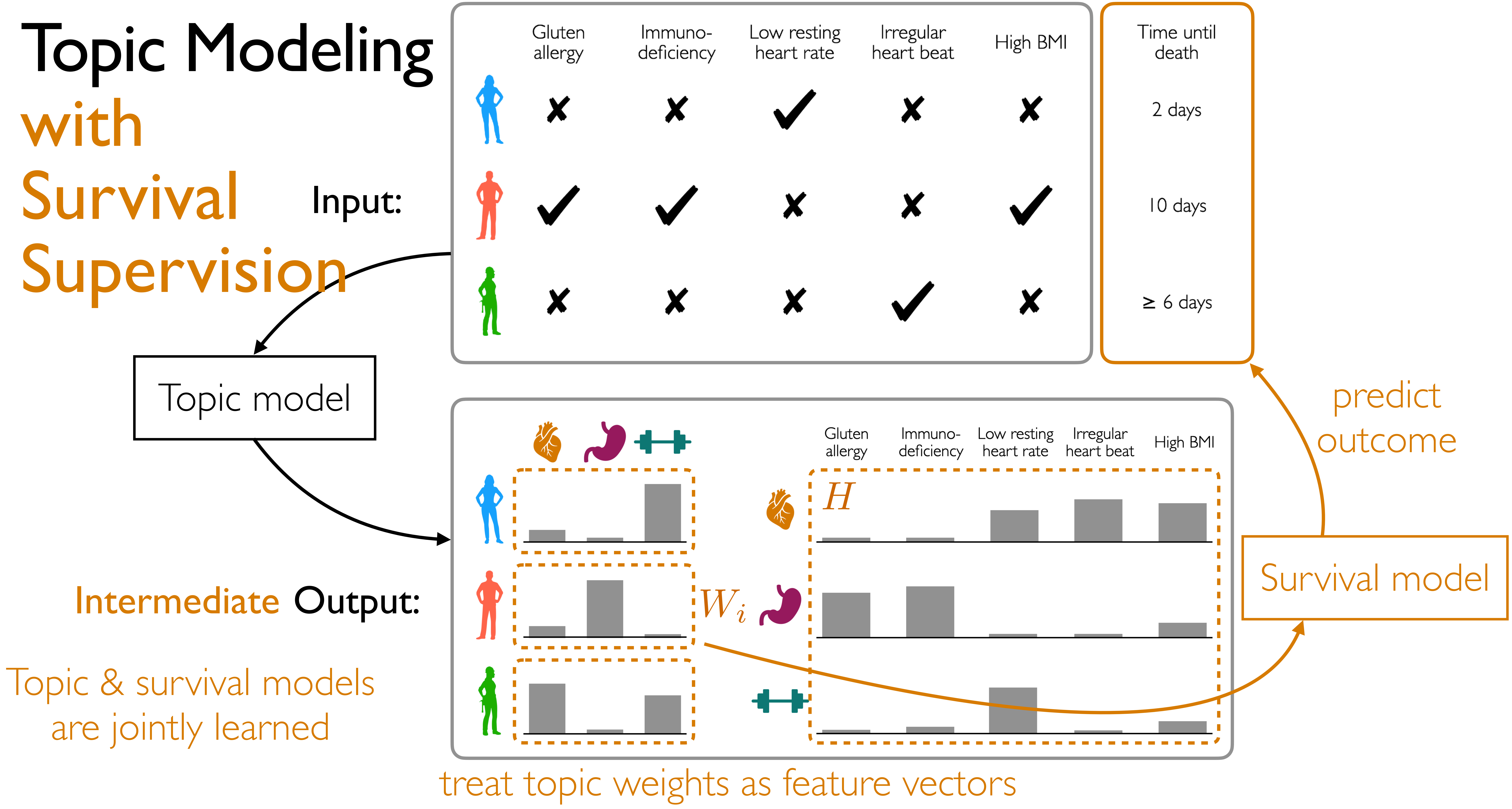


Topics encode how likely clinical measurements (“words”) are



Each subject has different amounts of different topics


Topic Modeling
Input:
Gluten allergy
Immuno-deficiency
Low resting heart rate
Irregular heart beat
High BMI
Topic model
Output:
Gluten allergy
Immuno-deficiency
Low resting heart rate
Irregular heart beat
High BMI
Standard topic modeling approaches are unsupervised


Topic Modeling with Survival Supervision
Input:
Gluten allergy
Immuno-deficiency
Low resting heart rate
Irregular heart beat
High BMI
Time until death
2 days
10 days
≥ 6 days
Topic model
Intermediate Output:
H
W_i
Survival model
predict outcome
Topic & survival models are jointly learned
treat topic weights as feature vectors

Neural Topic Modeling with Survival Supervision

For simplicity, we focus on LDA + Cox, producing a neural network variant of the approach by Dawson & Kendziorski (2012)

LDA neural approx. (Srivastava & Sutton 2017)

For training subject $i = 1, 2, \ldots, n$:

(a) Sample $\widetilde{W}_i \sim$ **LogisticNormal**$(0,$ **diag**$((k-1)/\alpha k))$

(b) Set topic weights to be $W_i =$ **softmax**$(\widetilde{W}_i)$

(c) Sample each "word" from **Categorical**(**softmax**$(W_i^\top H)$)

k = # of topics

α = Dirichlet prior hyperparameter

$H \in \mathbb{R}^{k \times d}$ topics' word distributions (unnormalized)

(d) Set the Cox partial log hazard output to be $\beta^\top W_i$

β = Cox regression coefficients

Survival supervision (Faraggi & Simon 1995, Katzman et al 2018)

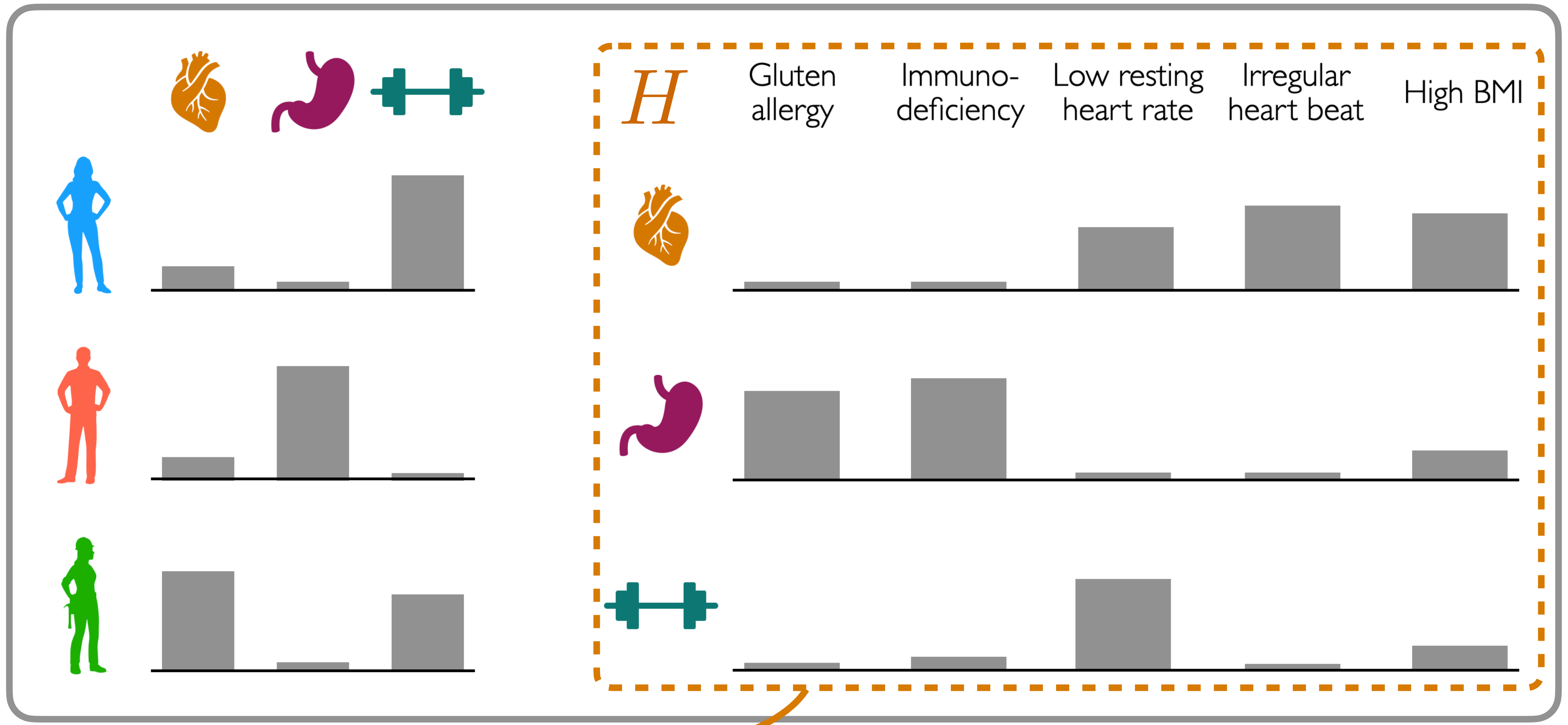
Loss function: LDA variational bound + $\boxed{\eta}$ Cox partial likelihood loss

hyperparameter

Implementation: we modify the software package *Scholar* (Card et al 2019) to obtain our approach *SurvScholar*

Model Interpretation

Intermediate Output:



After learning the model:

- Can interpret topics learned by looking at "top words"

We rank words by *relative* frequency (multiplicative factor compared to background frequencies)

Ranking by absolute frequencies not as interpretable due to common background words

- Each topic is associated with a Cox regression β coefficient

A topic having higher β coefficient $\Longrightarrow$ shorter survival time

- For any test subject, we can readily figure out the subject's topic weights

Datasets

Outcome: time until death

Dataset	Description	# subjects	# features	% censored
SUPPORT (Knaus et al 1995) split into 4 datasets corresponding to different diseases	1: acute respiratory failure/multiple organ system failure	4194	14	35.6%
	2: COPD/congestive heart failure/ cirrhosis	2804	14	38.8%
	3: cancer	1340	13	11.3%
	4: coma	591	14	18.6%
UNOS (unos.org/data)	heart transplant	62644	49	50.2%
METABRIC (Curtis et al 2012)	breast cancer	1981	24	55.2%
MIMIC (ICH) (Johnson et al 2016)	intracerebral hemorrhage	1010	1157	0%

Outcome: ICU length of stay

Experimental Setup

For all methods tested:

- Use 5-fold cross-validation on training data to select best hyperparameters
- With best hyperparameters, train on complete training dataset
- Evaluate performance on test data

Performance metric: time-dependent concordance index (Antolini et al 2005)

- Generalization of "area under the ROC curve" for survival analysis
  (value from 0 to 1 where 1 is perfect accuracy)

Specifically for our method *SurvScholar*:

- During cross-validation, pick model with fewest number of topics that has cross-validation concordance index within 0.005 of optimal

Intentionally favor *fewer* number of topics to make interpretability easier

Accuracy Benchmark

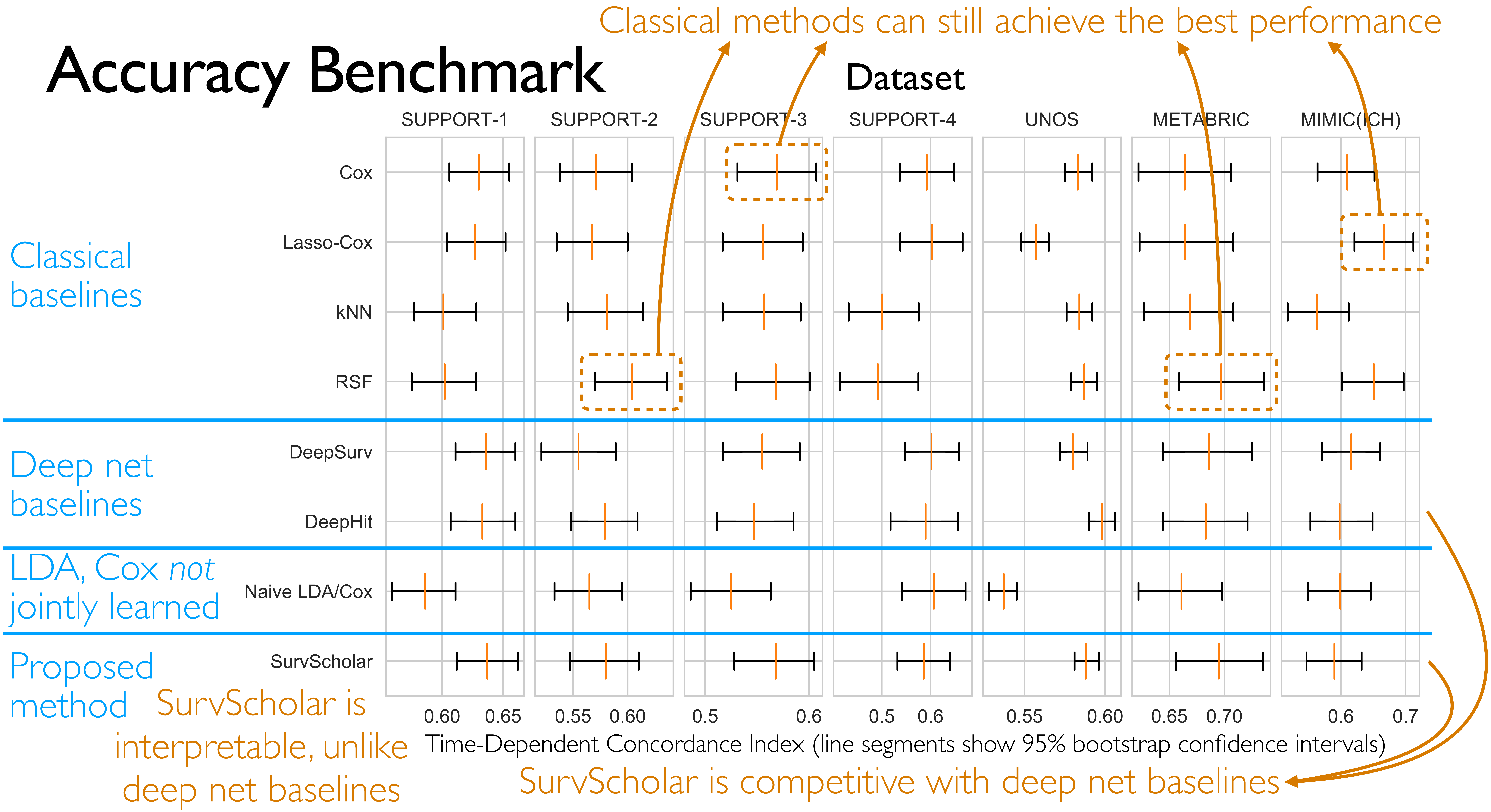


Illustration of Model Interpretation

Dataset: SUPPORT (cancer cohort)

One topic associated with shorter survival times

- hypotension
- hyponatremia
- multicomorbidity
- old age

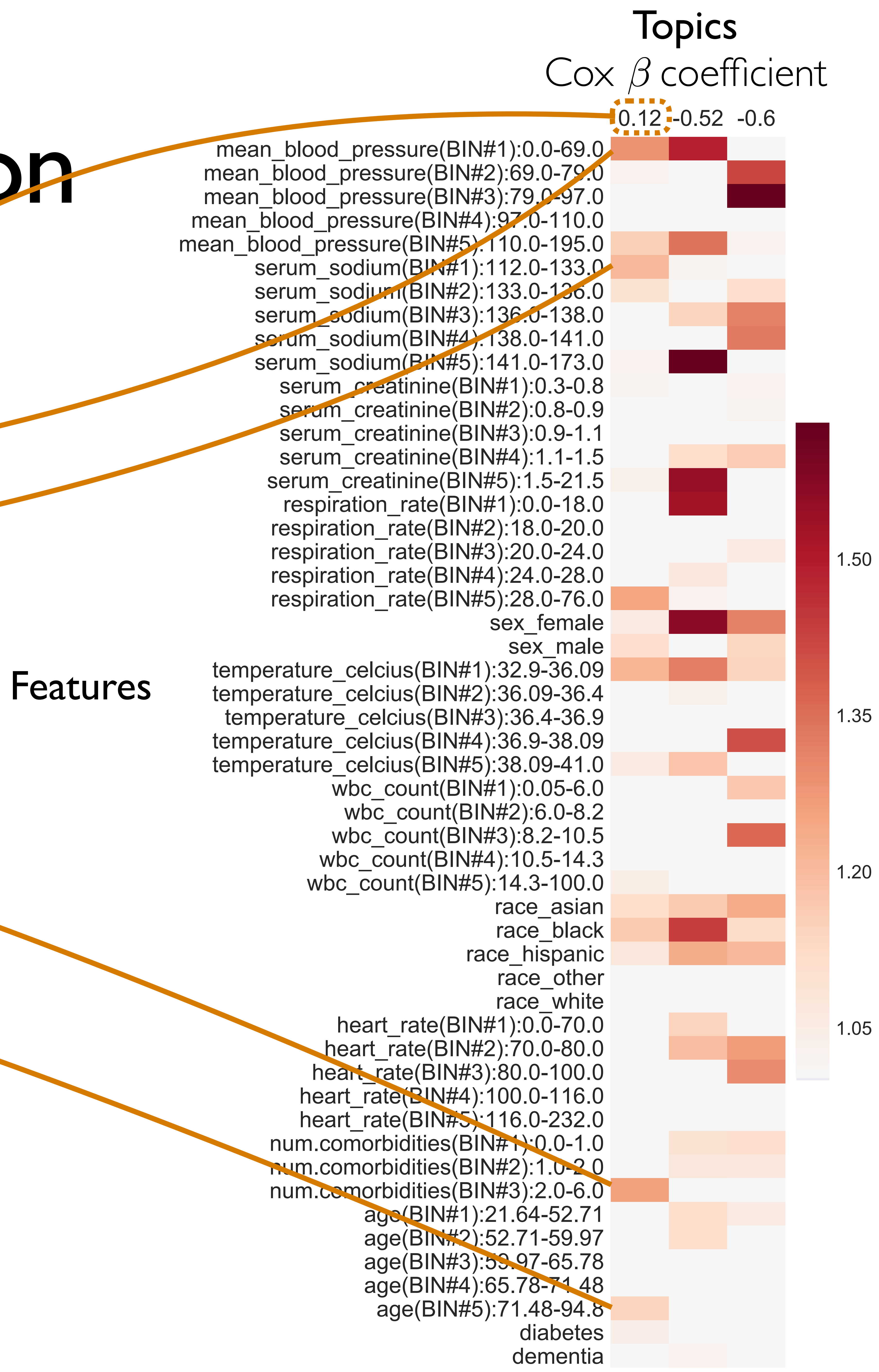


Illustration of Model Interpretation

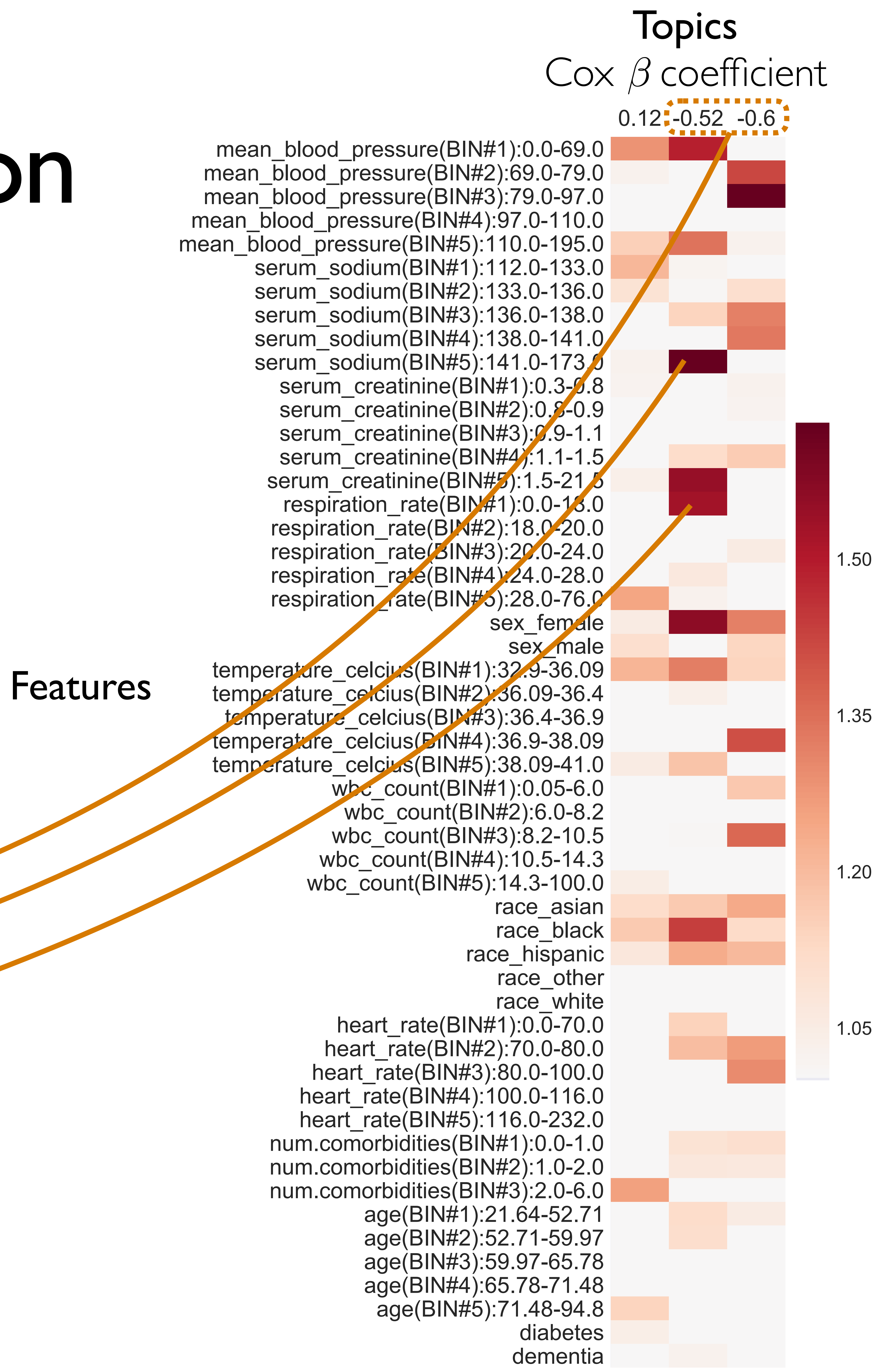
Dataset: SUPPORT (cancer cohort)

One topic associated with shorter survival times

- hypotension
- hyponatremia
- multicomorbidity
- old age

Two topics associated with longer survival times

- one topic has vital sign & laboratory derangements of sodium and creatinine
- other topic has normal vital sign & laboratory measurements



Discussion

Main contribution: neural net framework that combines topic modeling with survival analysis

Just need topic and survival models to have neural net formulations:

- The Scholar software package (Card et al 2019) we modify supports other topic models, e.g., SAGE (Eisenstein et al 2011), correlated topic models (Blei & Lafferty 2006)
- Can swap out Scholar altogether and use other neural topic models such as the Embedded Topic Model (Dieng et al 2019)
- Other survival models: Weibull accelerated failure time (Kalbfleisch & Prentice 2002), and any deep-learning-based survival model (Katzman et al 2018, Lee et al 2018, …)

How interpretable the final joint model is depends on the topic and survival models used!

When prediction accuracy is low, can look at topics learned to help debug

Still need to explore more topic/survival model combinations to see what works well

When # of features is very large, LDA struggles to identify most salient features