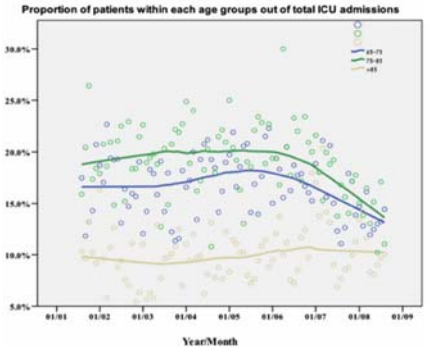
Possible Data Sets!

Jeremy Weiss, MD-PhD
Assistant Professor of Health Informatics

The latest version of MIMIC is MIMIC-III, which comprises over 58,000 hospital admissions for 38,645 adults and 7,875 neonates. The data spans June 2001 - October 2012. The



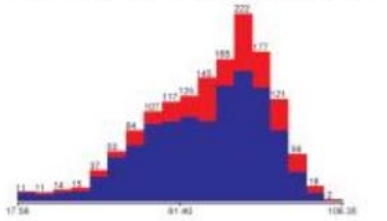
MIMIC II Deceased

(or MIMIC III if you need more data, but, more involved access process)

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

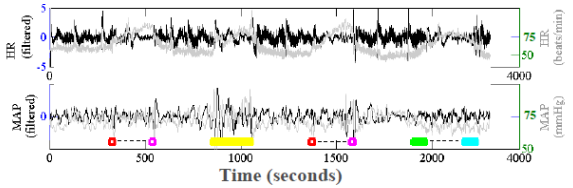
Leo Anthony G. Celi, MD, MS, MPH,[1,*] Robin J. Tang, MBA,[2] Mauricio C. Villarroel,[3] Guido A. Davidzon, MD, MS,[4] William T. Lester, MD,[5] and Henry C. Chueh, MD, MS[5]

(a) Age distribution among survivors and

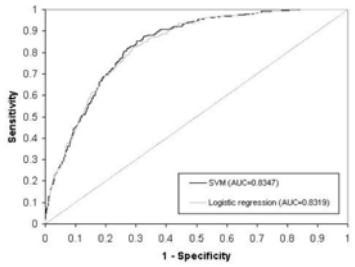


Red:
Mean (survivors) = 68.0
Standard deviation (survivors) = 16.0

Learning Outcome-Discriminative Dynamics in Multivariate Physiological Cohort Time Series



(b) Tilt-table experiment



Continuous NHANES

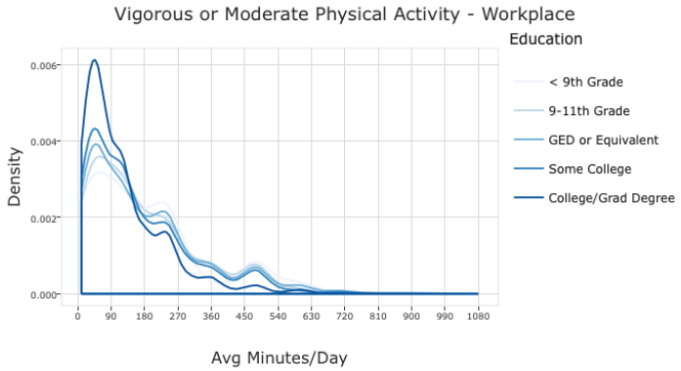
Cycles	1999-2000	2001-2002	2003-2004	etc...

Component	Demographics	Examination	Laboratory	Questionnaire	Dietary
Component Data Files	Demographics (including survey design variables)	Audiometry Blood Pressure Body Measures Muscular Strength Oral health Vision Exam etc...	Urine Collection Hepatitis HIV Heavy metals Plasma Glucose Total Cholesterol Triglycerides etc...	Alcohol use Balance Blood Pressure Diabetes Drug Use Social Support Vision Weight History etc...	Dietary Interview Supplement Use etc...

NHANES – National Health and Nutrition Examination Survey

library(RNHANES)

NHANES variables selected from feature subset selection using the Lasso	
Age	Poverty income ratio
Hemoglobin	On diabetes pills
On non drug diabetes interventions	On hypertension pills
Systolic blood pressure	Diastolic blood pressure
Education	Marital status



Other possible data sets:

Longitudinal Study of Aging:
https://www.cdc.gov/nchs/lsoa/index.htm

Amyotrophic Lateral Sclerosis:
https://nctu.partners.org/ProACT/Document/DisplayLatest/2

Head injury – CRASH I and CRASH II trials:
https://ctu-app.lshtm.ac.uk/freebird/index.php/available-trials/

More:
Imaging - https://sites.google.com/site/aacruzr/image-datasets
CDC - https://www.cdc.gov/nchs/data_access/ftp_data.htm

(health data you have access to, with permission of faculty mentor/research collaborators)

Exemplary Analyses

Example 1:

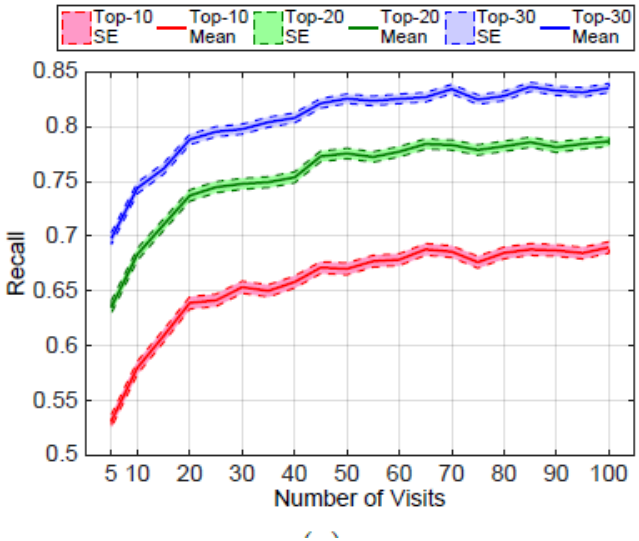
https://arxiv.org/abs/1511.05942

Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, Jimeng Sun

(Submitted on 18 Nov 2015 (v1), last revised 28 Sep 2016 (this version, v11))

Algorithms	Dx Only Recall @k		
	$k = 10$	$k = 20$	$k = 30$
Last visit	29.17		
Most freq.	56.63	67.39	71.68
Logistic	43.24	54.04	60.76
MLP	46.66	57.38	64.03
RNN-1	63.12	73.11	78.49
RNN-2	63.32	73.32	78.71
RNN-1-IR	63.24	73.33	78.73
RNN-2-IR	**64.30**	**74.31**	**79.58**



Example 2:

http://stm.sciencemag.org/content/7/299/299ra122.short

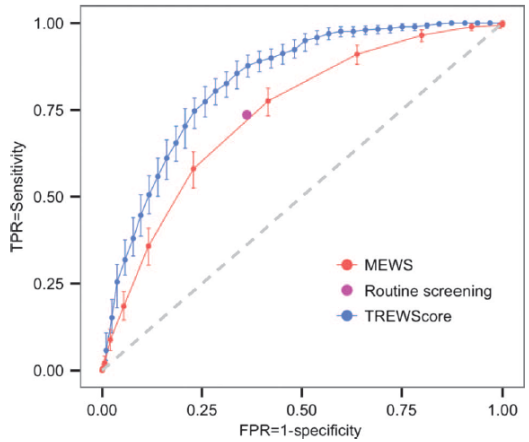
SEPSIS

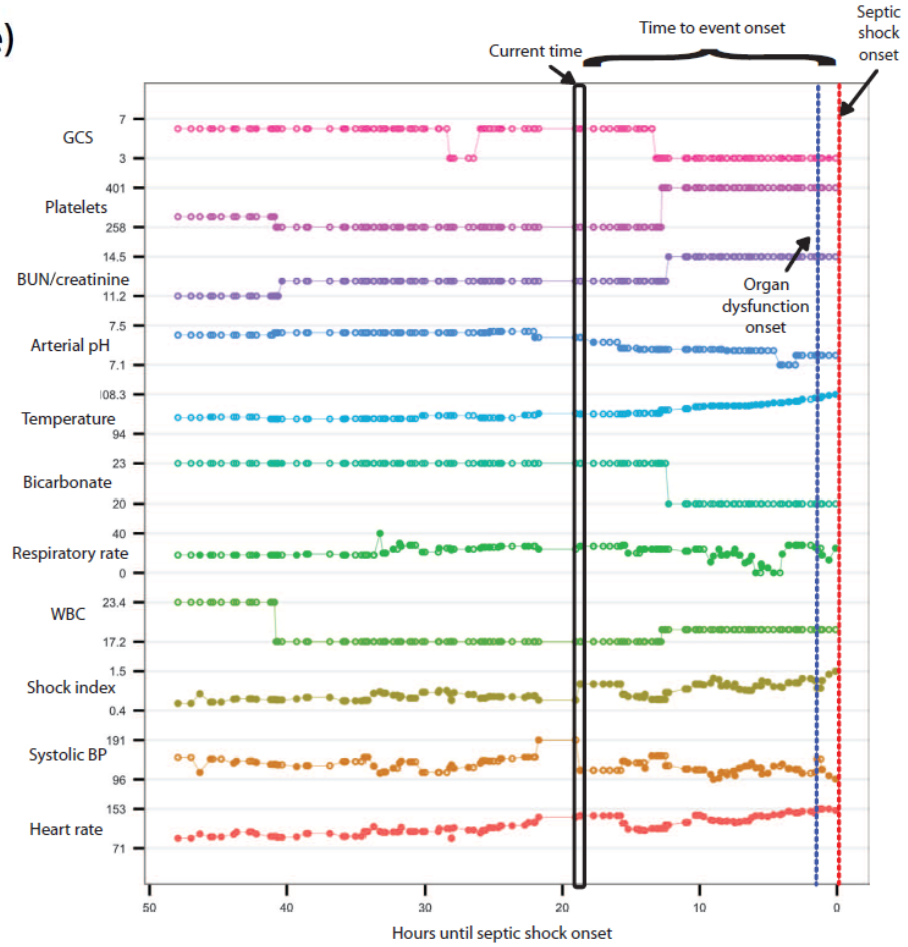
A targeted real-time early warning score (TREWScore) for septic shock

Katharine E. Henry,[1] David N. Hager,[2] Peter J. Pronovost,[3,4,5] Suchi Saria[1,3,5,6]*

Model development: Estimating model coefficients

To develop a model for predicting an individual's risk of developing septic shock, we fit a Cox proportional hazards model using the time until the onset of septic shock as the supervisory signal. Intuitively, this

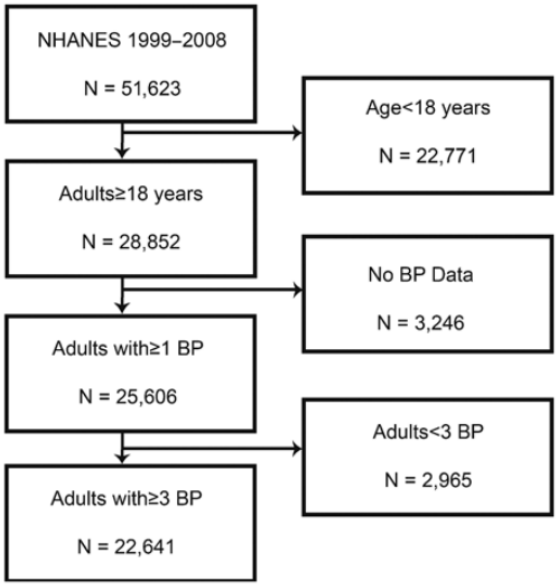


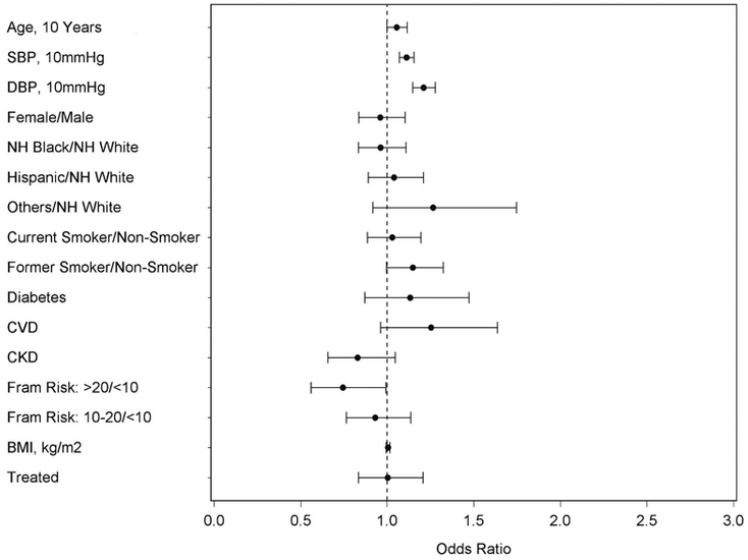


Example 3

Impact of the Number of Blood Pressure Measurements on Blood Pressure Classification in US Adults: NHANES 1999–2008

Joel Handler MD, Yumin Zhao PhD, Brent M. Egan MD





In contrast, patients with an initial BP above normal are often reclassified to a lower category, which supports recommendations for additional measurements.

Example 4

http://jamanetwork.com/journals/jamacardiology/article-abstract/2572174

Brief Report

February 2017

Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure

Comparison of Machine Learning and Other Statistical Approaches

Jarrod D. Frizzell, MD, MS[1]; Li Liang, PhD[2]; Phillip J. Schulte, PhD[3]; et al

Conclusions and Relevance Use of a number of ML algorithms did not improve prediction of 30-day heart failure readmissions compared with more traditional prediction models. Although there will likely be further applications of ML approaches in prognostic modeling, our study fits within the literature of limited predictive ability for heart failure readmissions.