Data, Decisions, and You:

Making Causality Useful and Usable in a Complex World

Samantha Kleinberg
Stevens Institute of Technology

CUMC Neuro-ICU


Photo by Chona Kasinger

COMMENTARY

A Framework for Reducing Alarm F

AUTHORS
Amogh Karnik, Bonafide, MD,
[1]The Commonw Pennsylvania; [2]Division of Ge Hospital of Phila [3]Center for Pe Children's Hosp Pennsylvania; [4]Perelman Sch of Pennsylvania [5]Leonard Davis University of Pen

KEY WORDS
alarm fatigue, c physiologic mo

ABBREVIATIC
ECG: electroca

www.hospitalpe
doi:10.1542/hp

Address corres P. Bonafide, MD of Philadelphia, Blvd, Suite 12N bonafide@ema

HOSPITAL PED

Anaesthesia, 2006, 6

Learnab equipme

P. M. Sanders

1 Professor of Cogni Factors; 2 PhD can St Lucia, Queensla

Summary
Melodic alarms tested for learna alarms over two than 30% of par Confusions pers (p = 0.011). Par medium priority alarms as soundi training identifie and found the t alarms are neede

AACN Advanced Critical Care
Volume 24, Number 4, pp.378-386

CNE

Alarm Fatigue

A Patient Safety Concern

Sue Sendelbach, RN, PhD, CCNS
Marjorie Funk, RN, PhD

ABSTRACT

Research has demonstrated that 72% to 99% of clinical alarms are false. The high number of false alarms has led to alarm fatigue. Alarm fatigue is sensory overload when clinicians are exposed to an excessive number of alarms, which can result in desensitization to alarms and missed alarms. Patient deaths have been attributed to alarm fatigue. Patient safety and regulatory agencies have focused on the issue of alarm fatigue, and it is a 2014 Joint Commission National Patient Safety Goal. Quality improvement projects have demonstrated that strategies such as daily electrocardiogram electrode changes, proper skin preparation, education, and customization of alarm parameters have been able to decrease the number of false alarms. These and other strategies need to be tested in rigorous clinical trials to determine whether they reduce alarm burden without compromising patient safety.

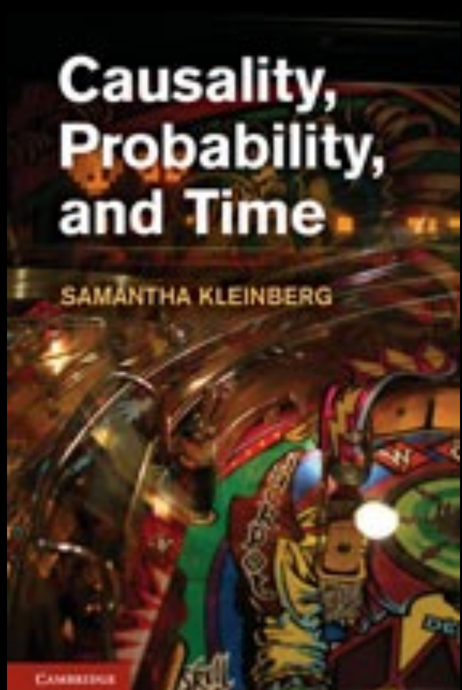
Keywords: alarm fatigue, patient safety, regulatory agencies

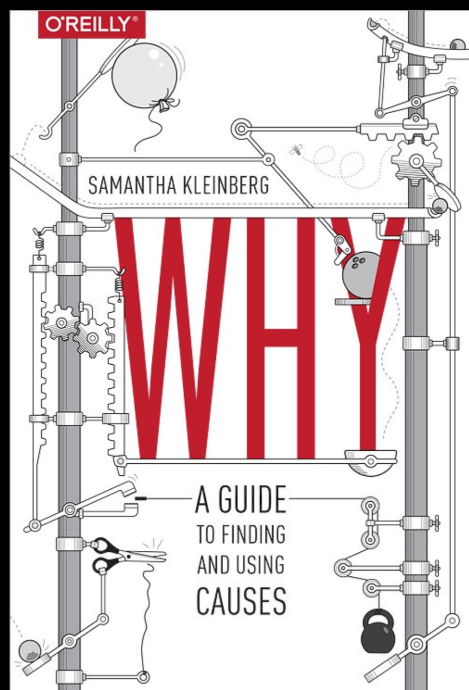
What doctors and patients actually want to know: why is this happening and what can they do about it?

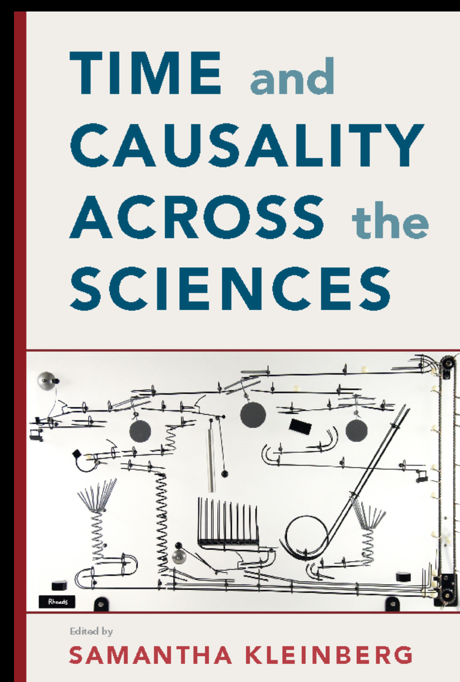
Data to knowledge

"The second lesson is that, when trying to make predictions you needn't worry too much about why your models work… And, in the prediction business, you just need to know that something works, not why."

Stephens-Davidowitz, S. (2017) Everybody Lies. Bloomsbury.

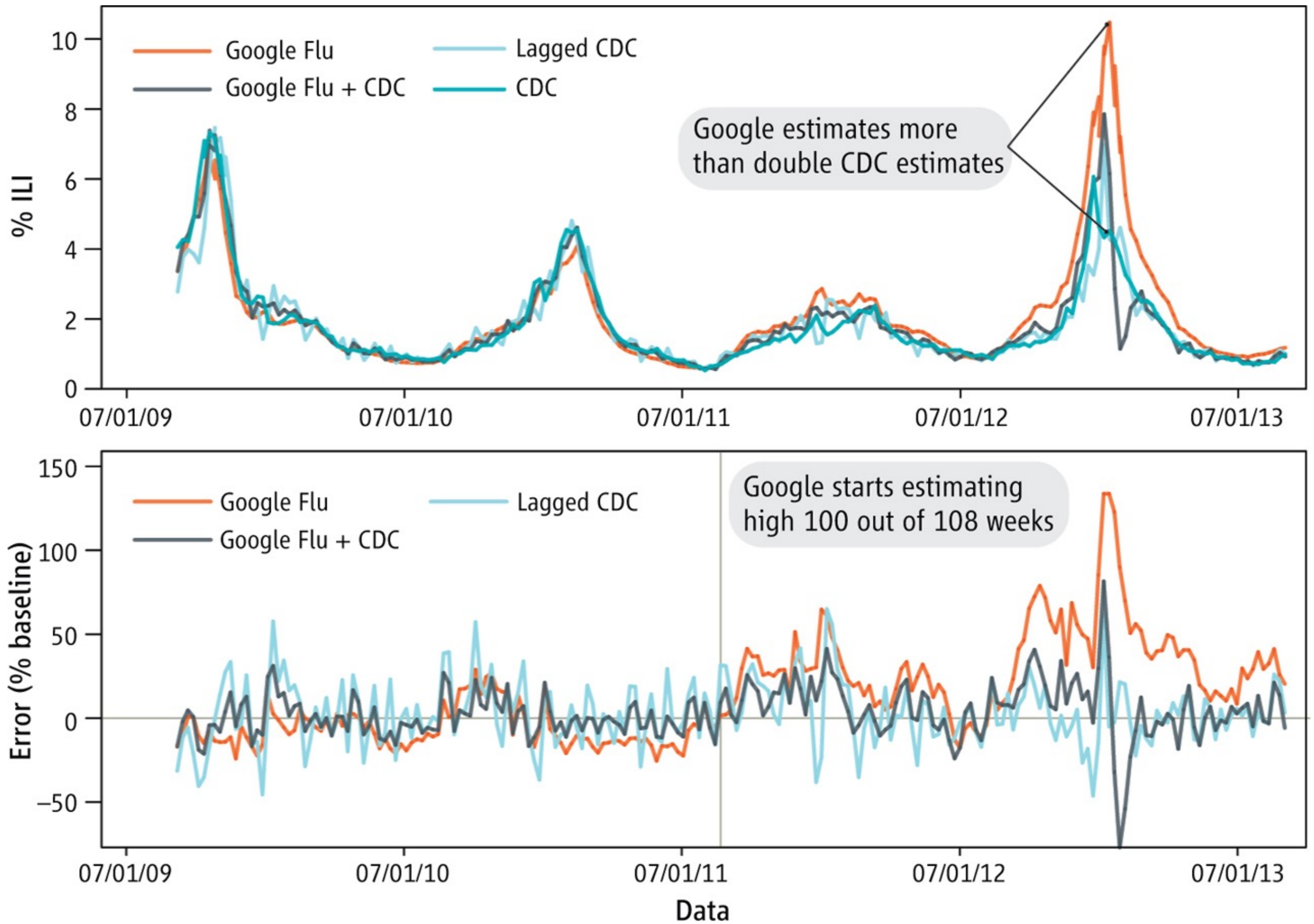






Action requires causality

Predict



David Lazer et al. Science 2014;343:1203-1205

Science AAAS

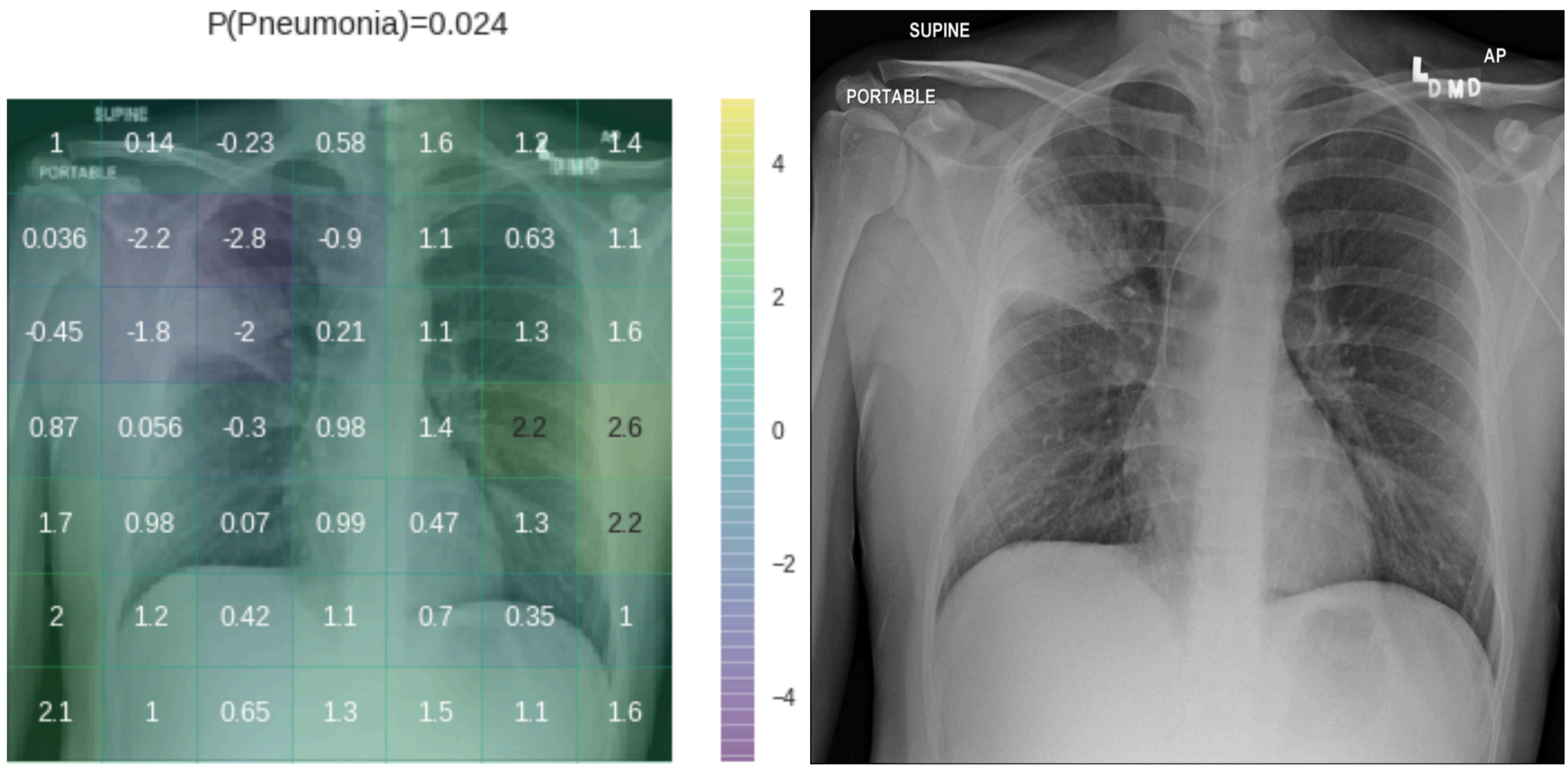
Published by AAAS

People should stop training radiologists now. It’s just *completely* obvious that within five years deep learning is going to do better than radiologists because it’s going to be able to get a lot more experience.

Geoff Hinton in 2016

https://www.youtube.com/watch?v=2HMPRXstSvQ

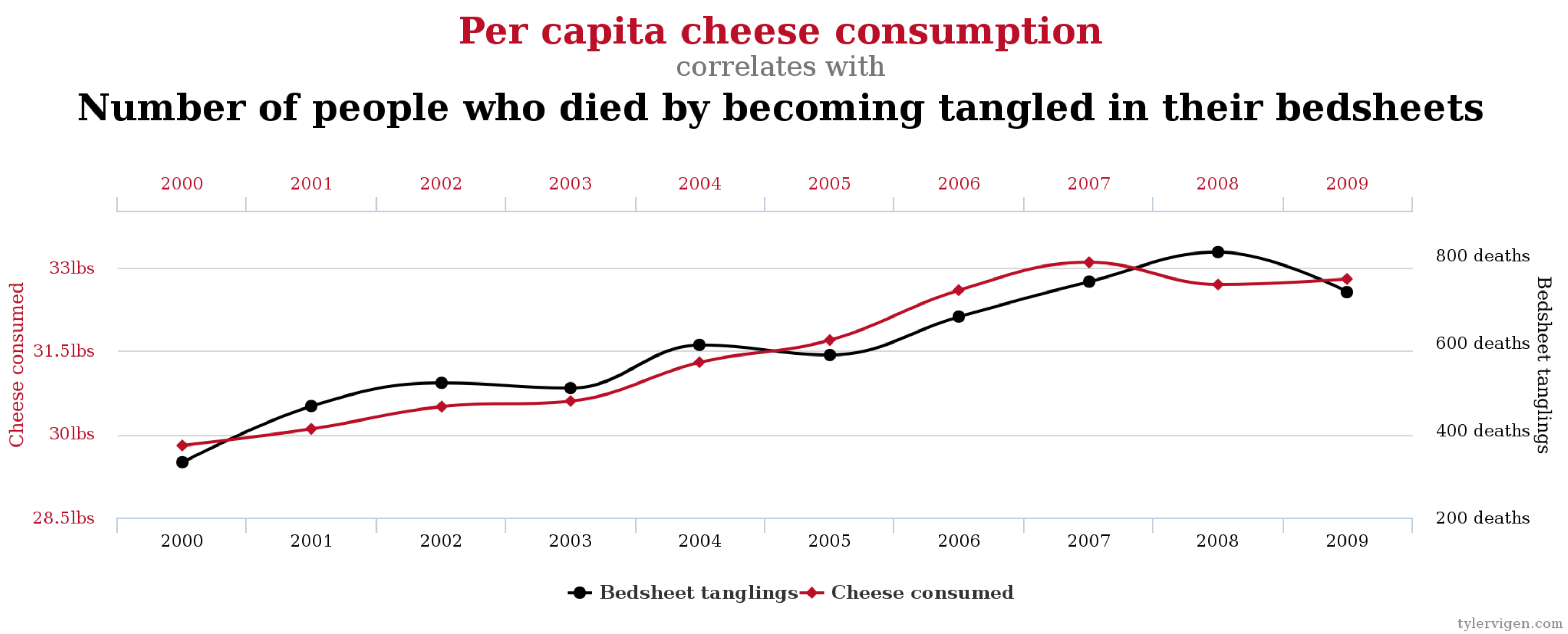
Predict



Patient with pneumonia: heatmap of CNN on left, original image on right

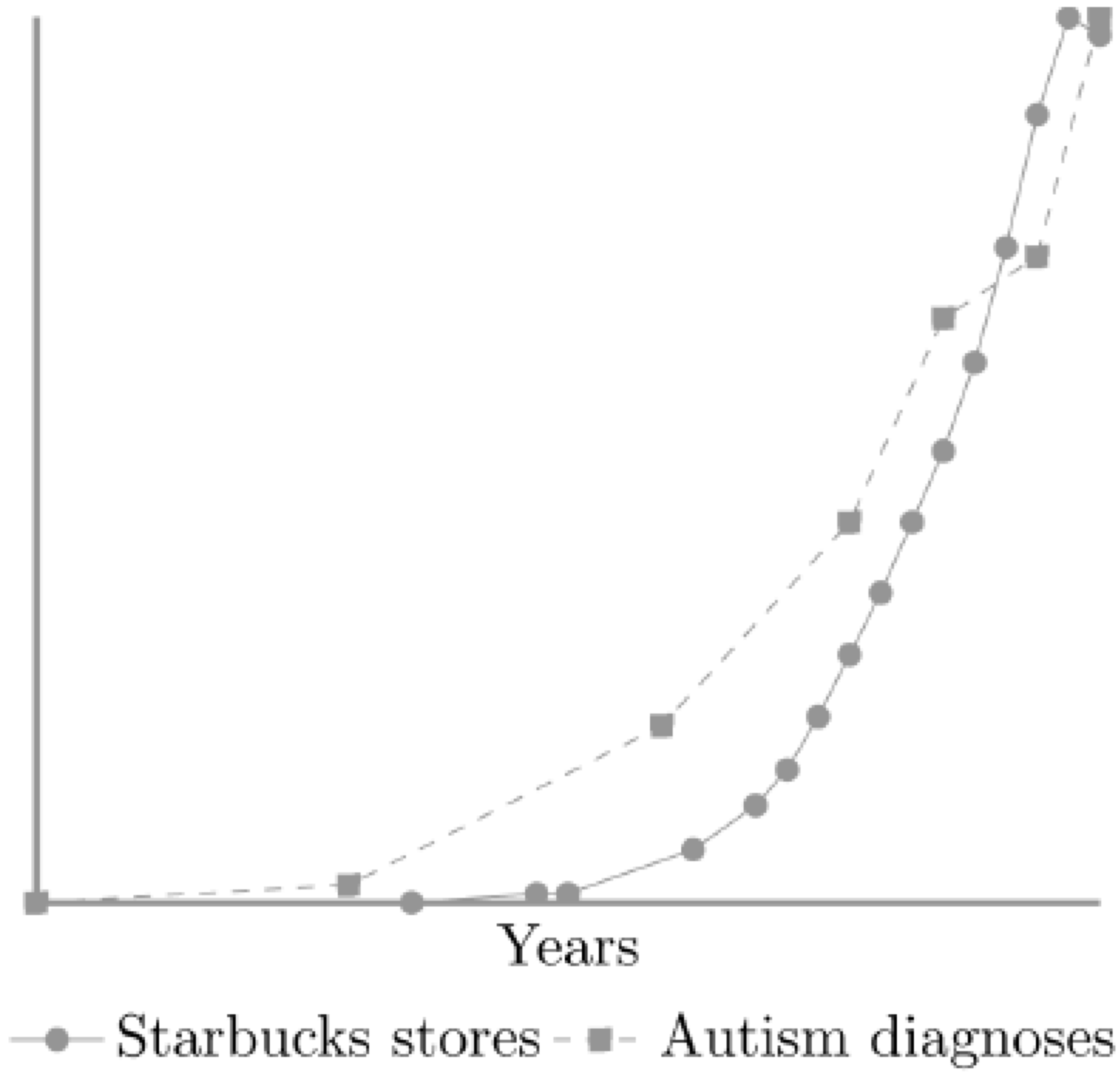
https://jrzech.medium.com/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98

Explain

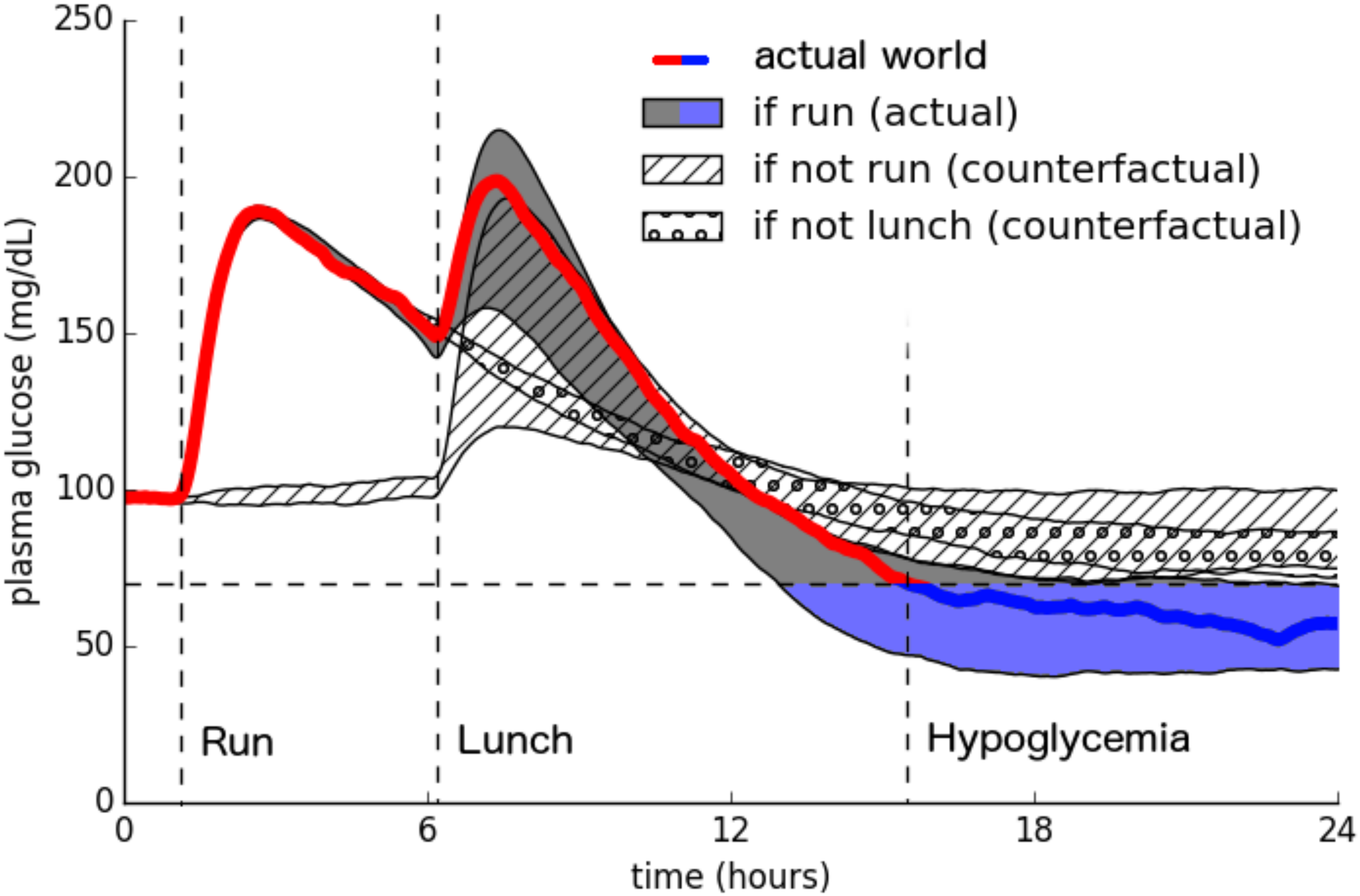


tylervigen.com

Explain



Explain



C. Merck and S. Kleinberg. Causal explanation under indeterminism: A sampling approach. AAAI, 2016.

Control

Reducing class sizes in Tennessee led to better scores on standardized tests

Will it work in California?

Bohrnstedt, G. W. and Stecher, B. M. (eds.) (2002). What We Have Learned about Class Size Reduction in California . American Institutes for Research, Palo Alto, CA.

Can we learn..

risk factors for heart failure

Kleinberg & Elhadad (2013) AMIA

what leads to an individual's hyperglycemic episodes

Heintzman & Kleinberg (2016) JBI

causes of secondary brain injury

Claassen et al. (2016) PLoS ONE

from observing people?

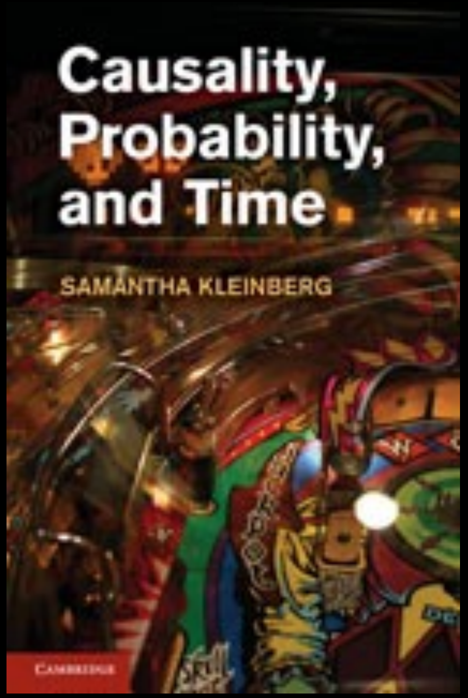
Logic-based causal inference

Complex, temporal relationships

$$v \leadsto^{\geq 15, \leq 40}_{\geq 0.4} g$$

(PCTL + some additions)

$$\varepsilon_{avg}(c,e) = \frac{P(e \mid c \wedge x) - P(e \mid \neg c \wedge x)}{|X \backslash c|}$$



Can we trust the data?

Application: learning from patient generated health data

- Chronic diseases are managed primarily by individuals

- Outpatient data brings even more uncertainty than EHR data

FDA guidance

95% of fingerstick BG values must be within 15% of the actual value

A value of 150 could be [128, 172]

A value of 70 could be [60,81]

CGM accuracy is also a function of time

Visit	Number of Readings / Unique Patients	Mean Absolute Relative Difference 95% Confidence Interval	Percentage of Readings within 15 mg/dL or 15%[a]
Day 1	2,665 / 35	11.6% (10.0, 13.1)	79%
Day 7	2,926 / 35	9.8% (7.9, 11.7)	86%

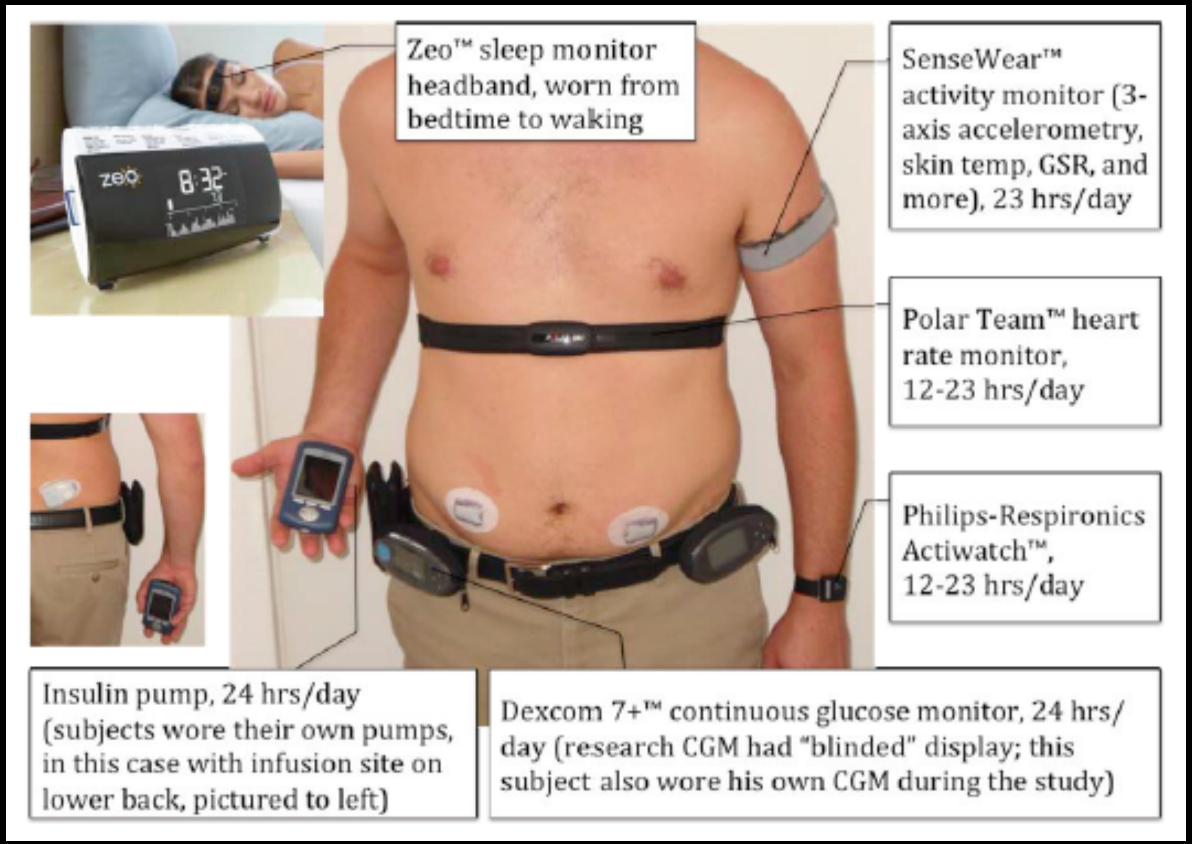
https://www.fda.gov/media/112142/download

17 subjects with T1DM, sensor data (collected for >72 hours)

Used causal inference methods + body-worn sensors to find cause of changes in glycemia

- intense activity leads to hyperglycemia in 5-30min

only found when modeling uncertainty

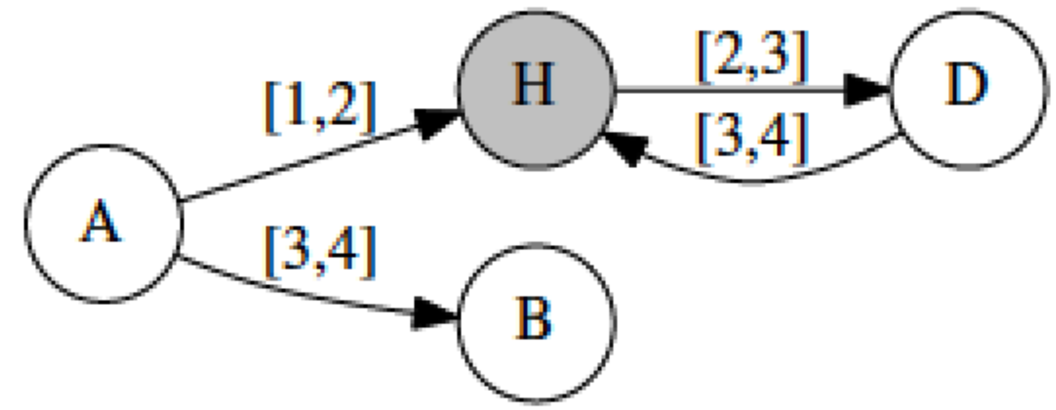


N. Heintzman and S. Kleinberg. Using Uncertain Data from Body-Worn Sensors to Gain Insight into Type 1 Diabetes. Journal of Biomedical Informatics (2016)

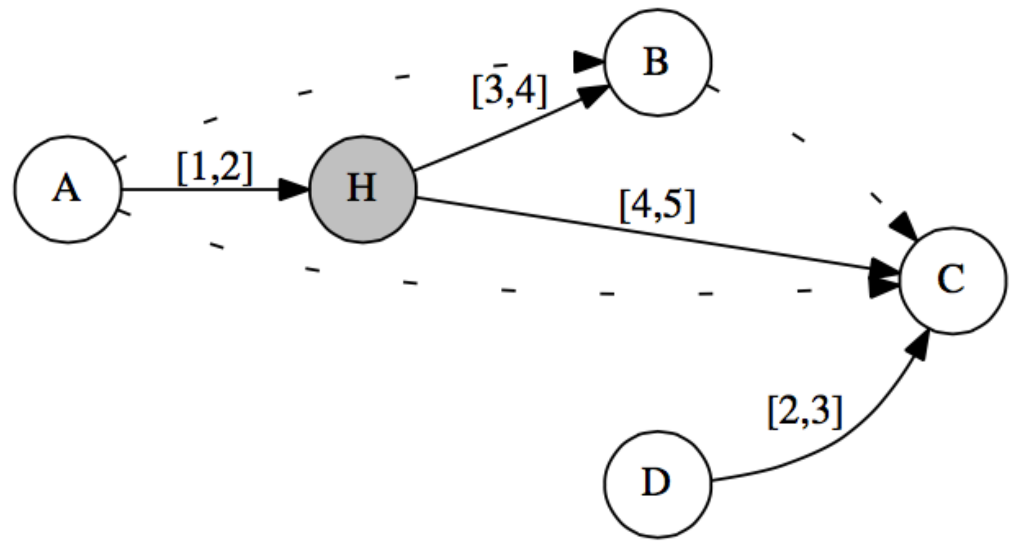
Latent variables are not always latent

- Leverage prior knowledge (experts, other experiments)
- Be robust to wrong/inapplicable knowledge
    - Reconstruct time series
    - Identify inconsistencies
    - Iterate

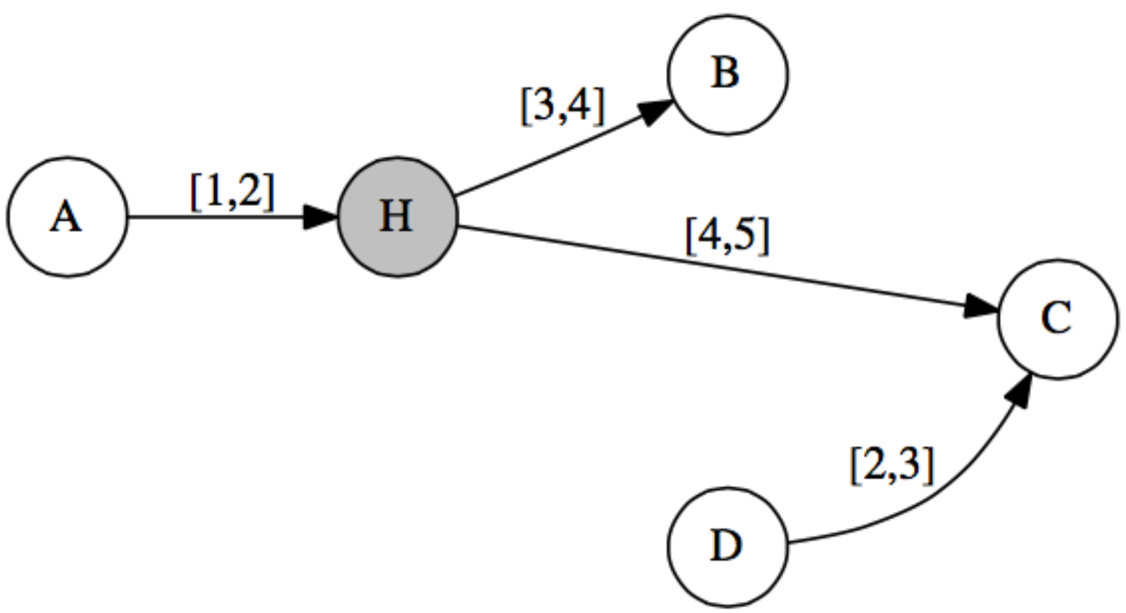
1. reconstruct data
2. find errors
3. infer causes



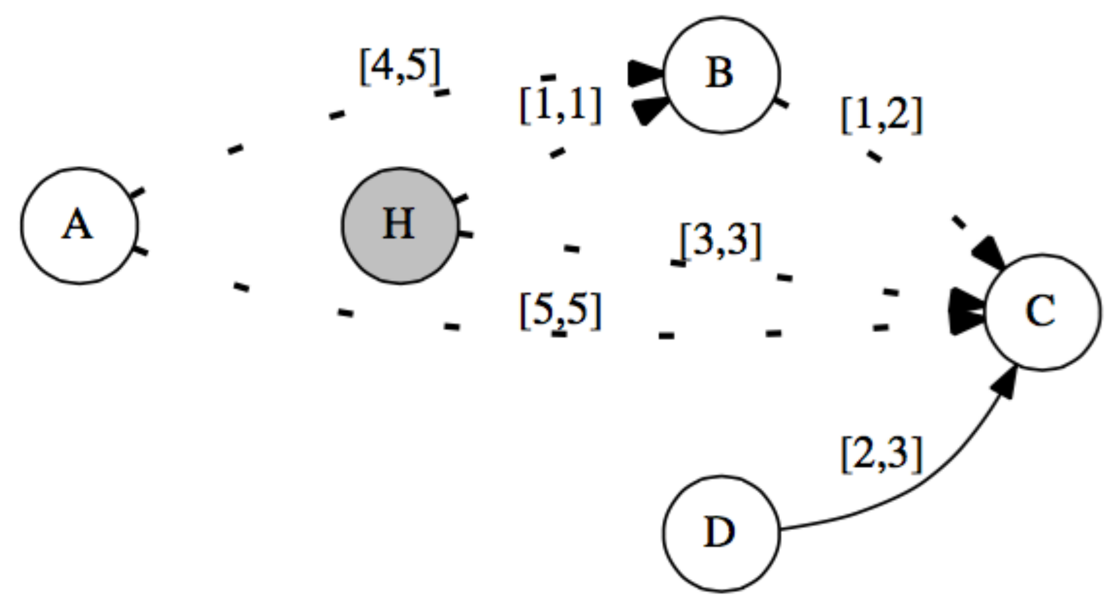
$$e = \begin{cases} Max(D[e,t-s]\ldots D[e,t-r]), & \text{if } e \in pa(v) \\ Max(D[e,t+r]\ldots D[e,t+s]), & \text{if } e \in ch(v) \\ Max(\bigcup_y Max(D[y,t+r]\ldots D[y,t+s])) \\ \text{where} \quad y \in (ch(v) \cap ch(e)), \text{and} \quad e \in pa(ch(v)) \end{cases}$$



truth



ours



tsFCI

We can use knowledge of the effect of meals on glucose to recover latent meals and their effects

We find: exercise causes meal in 60-85min, moderate exercise causes hypo in 70-90min, 67 meals recovered
tsFCI: 1 latent variable, and hypo/hyper cause themselves

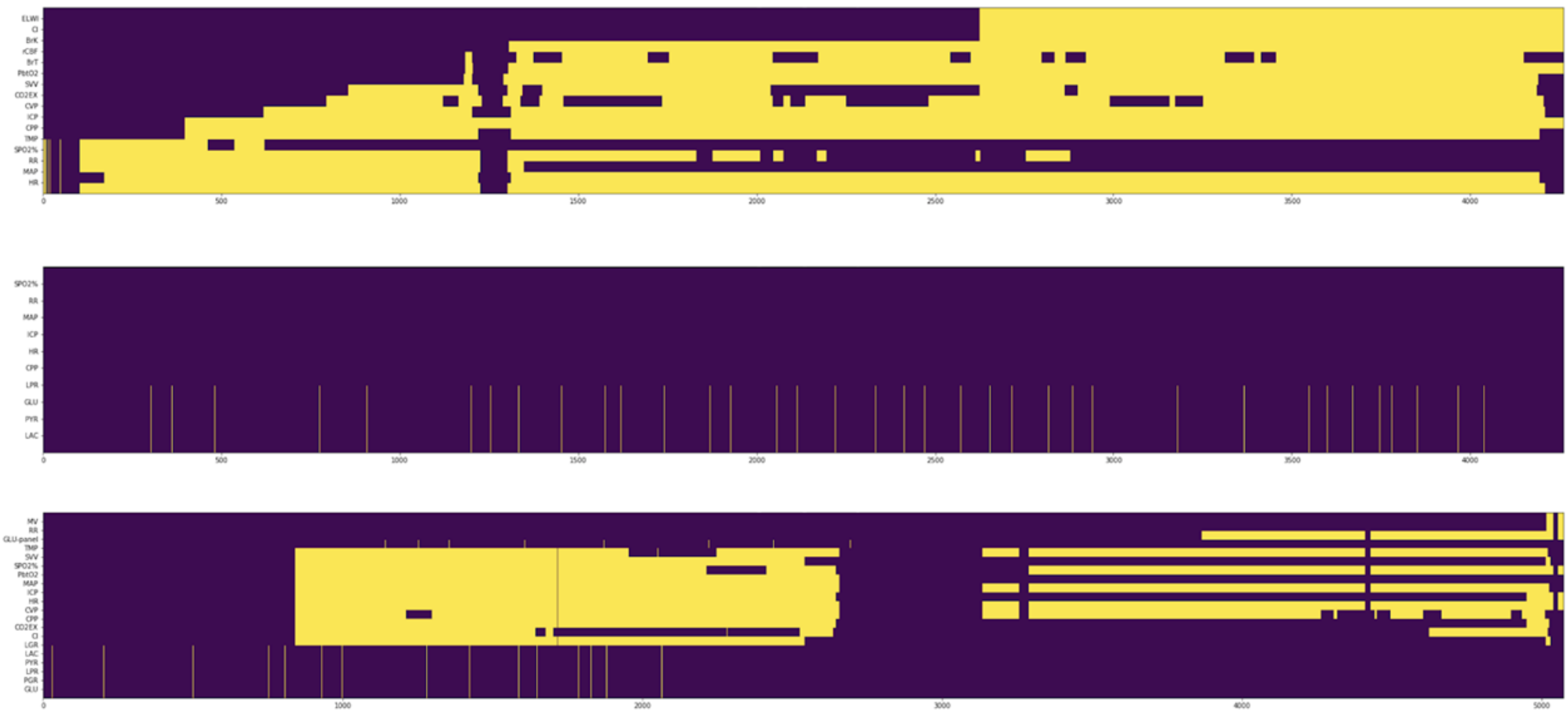
M. Zheng, and S. Kleinberg. (2019) Using Domain Knowledge to Overcome Latent Variables in Causal Inference from Time Series. Machine Learning for Healthcare.

Stroke

- 98 patients with subarachnoid hemorrhage

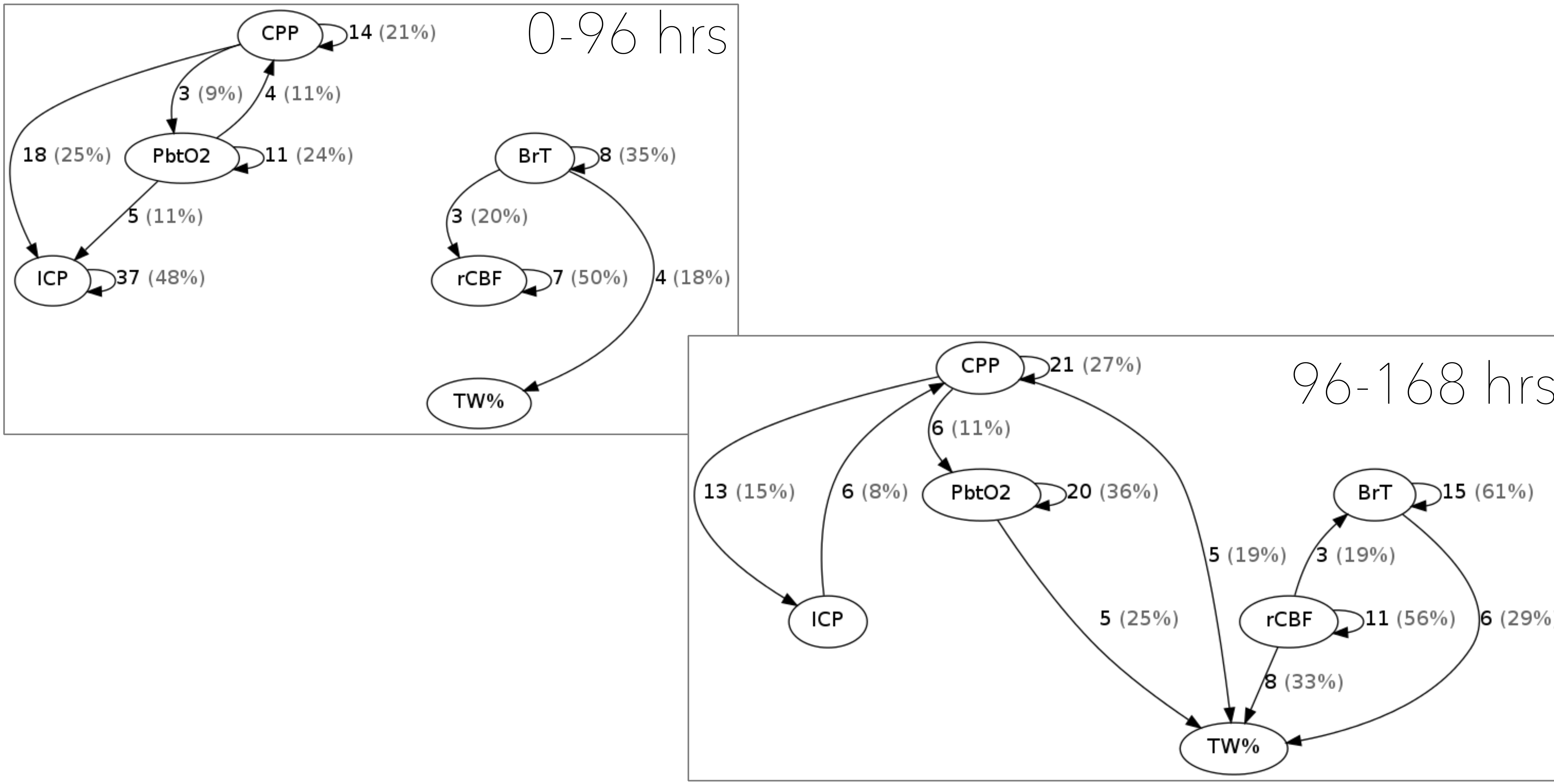
- Monitoring included
  - Depth and surface EEG
  - Microdialysis
  - Physiologic measurements

Data for 3 ICU patients. Purple = missing

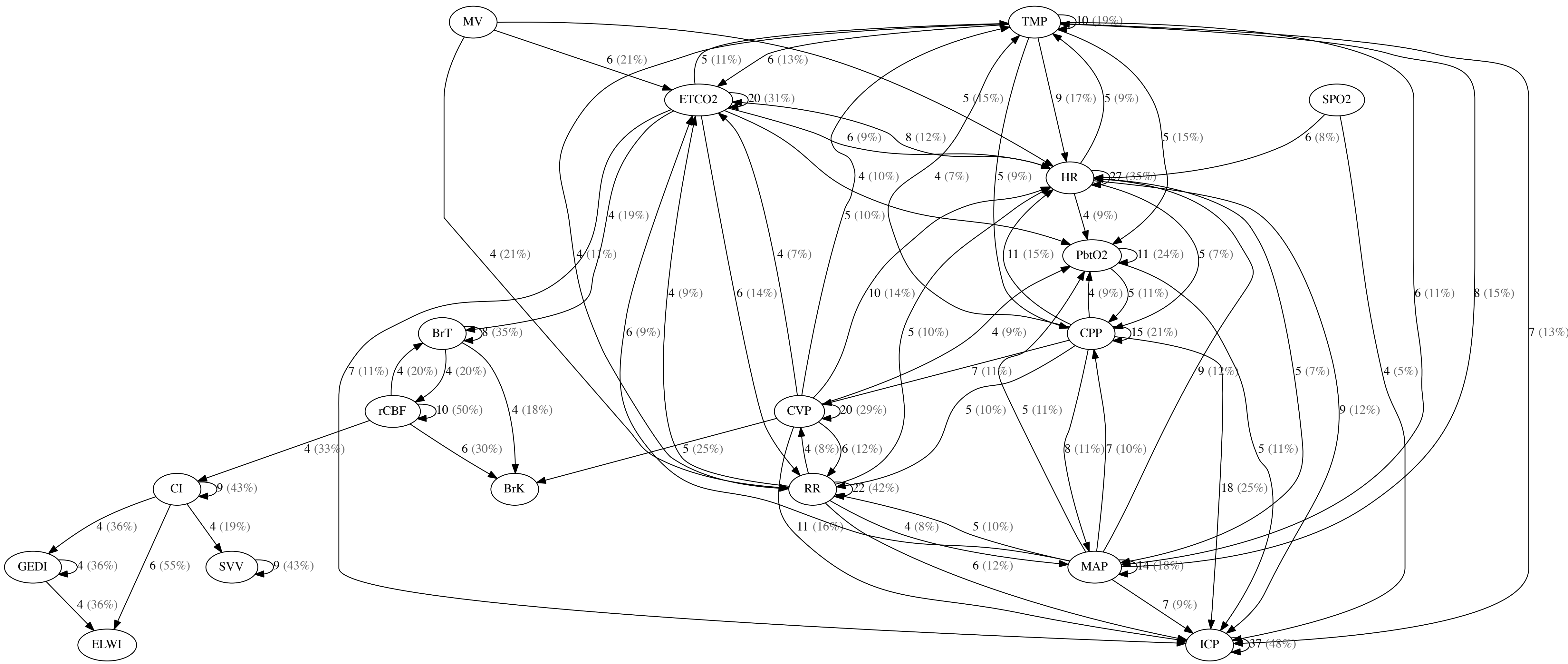


Time

S. A. Rahman, Y. Huang, J. Claassen, N. Heintzman, and S. Kleinberg. Combining Fourier and Lagged k-Nearest Neighbor Imputation for Biomedical Time Series Data. Journal of Biomedical Informatics (2015)

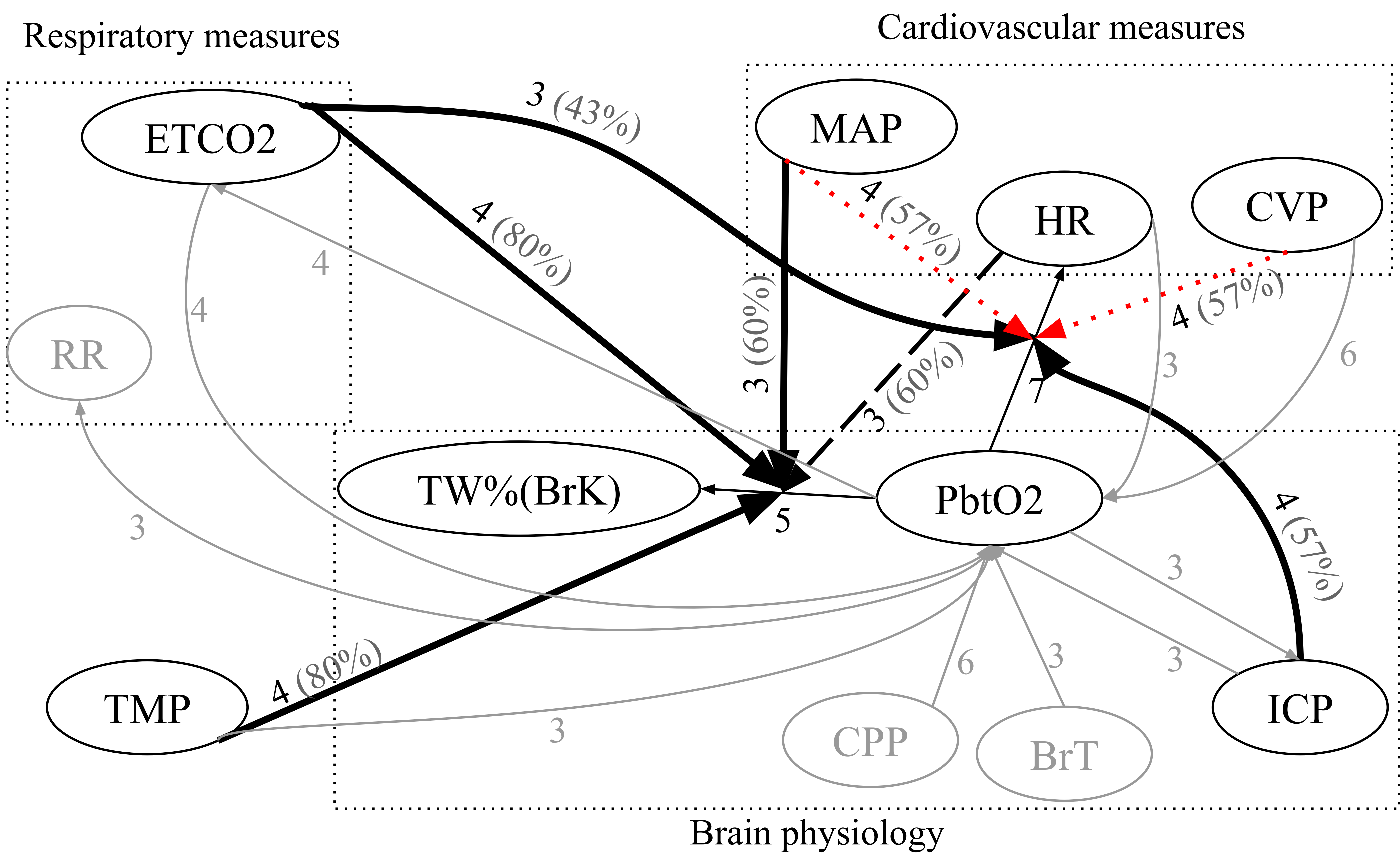


Claassen J, Rahman SA, Huang Y, Frey H, Schmidt M, Albers D, Falo CM, Park S, Agarwal S, Connolly ES, Kleinberg S (2016) Causal structure of brain physiology after brain injury. PLoS ONE.



Claassen J, Rahman SA, Huang Y, Frey H, Schmidt M, Albers D, Falo CM, Park S, Agarwal S, Connolly ES, Kleinberg S (2015) Causal structure of brain physiology after brain injury. PLoS ONE.

More descriptive structures



Zheng, M, Claassen J, Kleinberg S (2018) Automated Identification of Causal Moderators in Time-Series Data. ACM SIGKDD Workshop on Causal Discovery.

Does this actually help people make decisions?

From knowledge to action

Method	Negative Log Likelihood	p-value	MSE	p-value
Nile Data (200 Training Points, 462 Test Points)				
GPTS	1.19±0.0548	0.196	0.579±0.0976	0.356
♡ GPTS-CP	1.19±0.0548	0.167	0.583±0.0989	0.335
ARGP	1.18±0.0510	0.202	0.568±0.0940	0.410
♡ ARGP-CP	**1.15 ± 0.0555**	N/A	**0.553 ± 0.0962**	N/A
Kalman	1.17±0.0508	0.361	0.562±0.121	0.453
TIM	1.49±0.0714	<0.001	1.16±0.161	<0.001
♡ NSGP (grid)	1.15±0.0655	0.490	0.585±0.0988	0.321
Bee Waggle Dance Data (250 Training Points, 807 Test Points)				
GPTS	8.02±0.504	<0.001	8.44±0.745	<0.001
♡ GPTS-CP	4.54±0.188	<0.001	3.13±0.241	<0.001
ARGP	4.35±0.167	0.007	2.98±0.224	0.008
♡ ARGP-CP	**4.07 ± 0.150**	N/A	**2.62 ± 0.195**	N/A
Kalman	4.39±0.176	0.002	2.93±0.215	0.016
TIM	4.54±0.177	<0.001	3.25±0.237	<0.001
♡ NSGP (HMC)	4.19±0.212	<0.001	3.17±0.230	<0.001
Whistler Snowfall Data (500 Training Points, 13380 Test Points)				
GPTS	1.48±0.0455	<0.001	0.780±0.0333	<0.001
♡ GPTS-CP	1.17±0.0183	<0.001	0.689±0.0294	<0.001
ARGP	1.31±0.0395	<0.001	0.637±0.0268	0.143
♡ ARGP-CP	-0.604±0.0385	<0.001	0.750e±0.0315	<0.001
Kalman	1.28±0.0373	<0.001	**0.614 ± 0.0254**	0.589
TIM	1.47±0.0284	<0.001	1.01±0.0387	<0.001
♡ NSGP (grid)	**−1.98 ± 0.0561**	N/A	0.618±0.0242	N/A

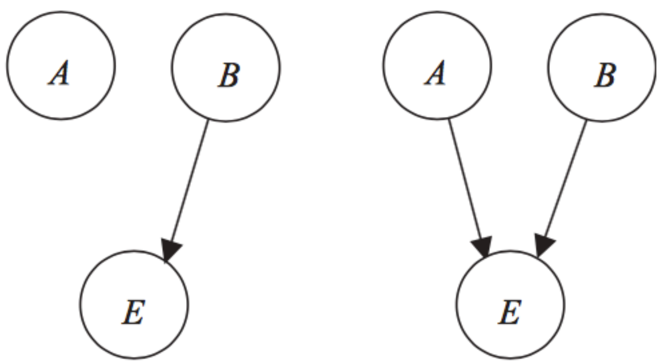
DAG	G-causal. linear	TiMINo linear	TS-LiNGAM
correct	13%	83%	19%
wrong	87%	7%	81%
no dec.	0%	10%	0%

Table 1: Exp.2: Gaussian data and linear instantaneous effects: only TiMINo mostly discovers the correct DAG.

Table 1: Results on simulated data. Run time is in seconds. *Run time for DBNs is a user-specified parameter.

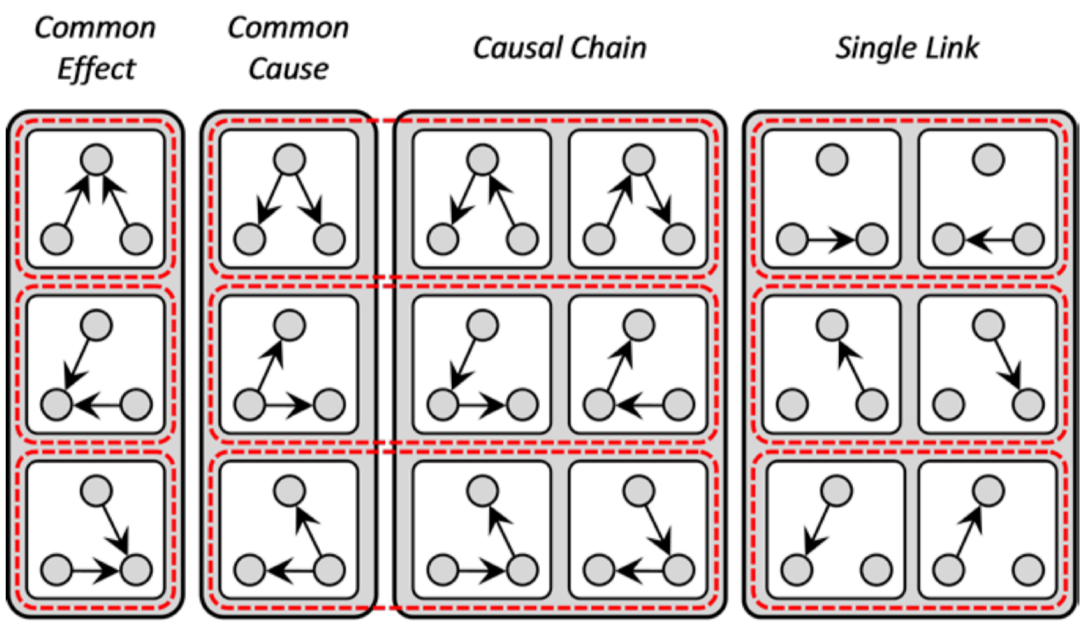
	Common cause & effect			Random			Finance		
Method	FDR	FNR	run time	FDR	FNR	run time	FDR	FNR	run time
DBNs	0.000	0.006	7200*	0.004	0.025	72000*	0.152	0.013	72000*
Granger	0.488	0.000	23	0.798	0.007	230	0.718	0.015	905
$\varepsilon_{avg}(c,e)$	0.650	0.000	1567	0.053	0.015	68186	0.078	0.012	85678
$\alpha(c,e)$	0.000	0.000	16	0.001	0.006	5088	0.036	0.006	8456

- Which causes do we find?
  - Stronger causes
  - Earlier causes
  - Modifiable factors
- Are some causes more valuable than others?



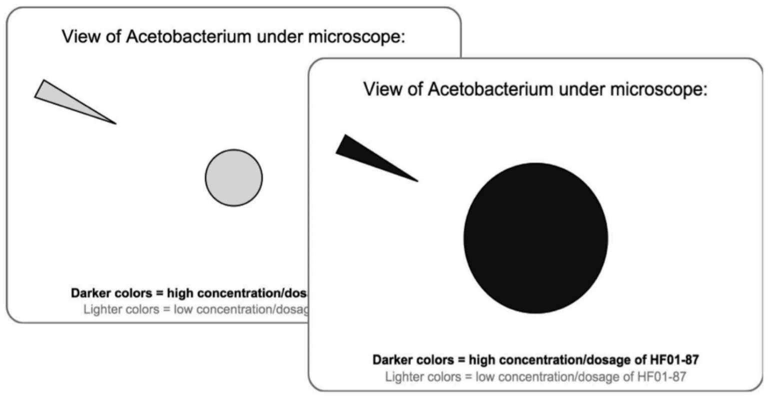
Super pencils/
Blicket detector

Griffiths et al. 2011



Mind-reading aliens

Mayrhofer & Waldmann 2011



Drug+microorganism size

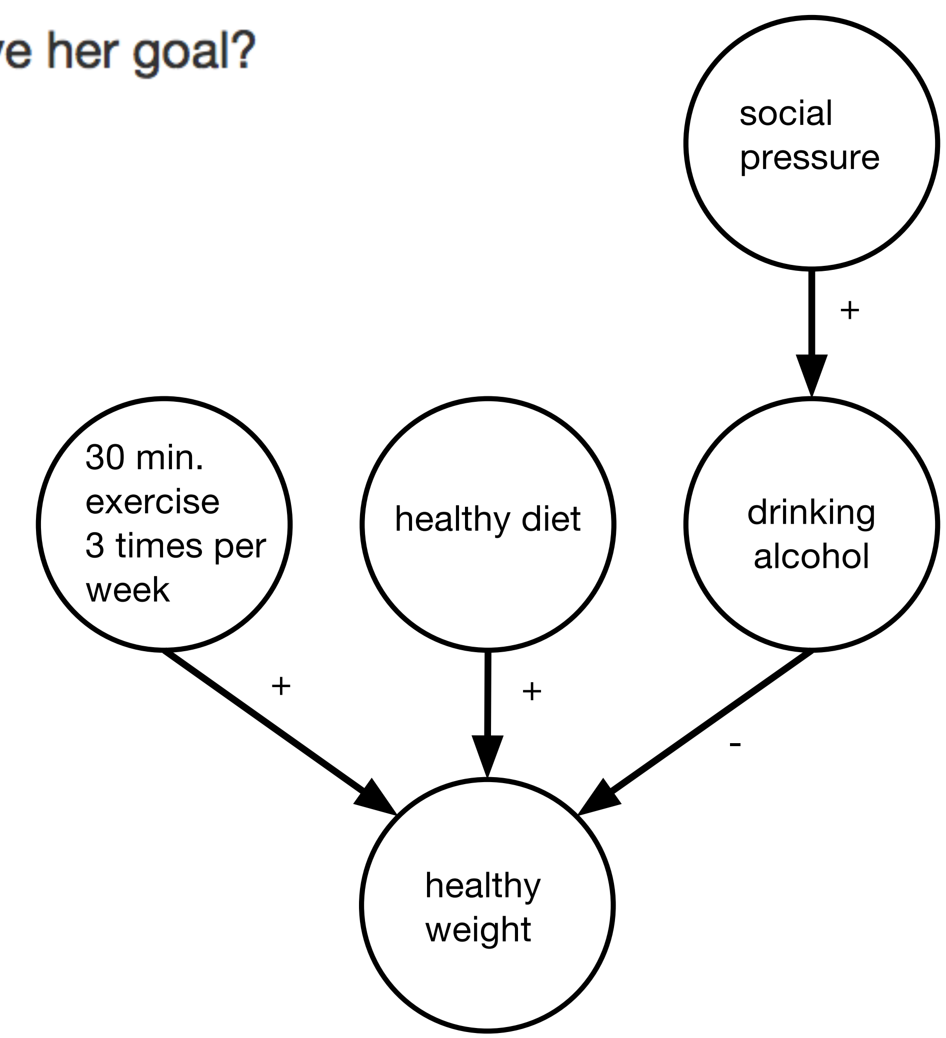
Soo & Rottman 2018

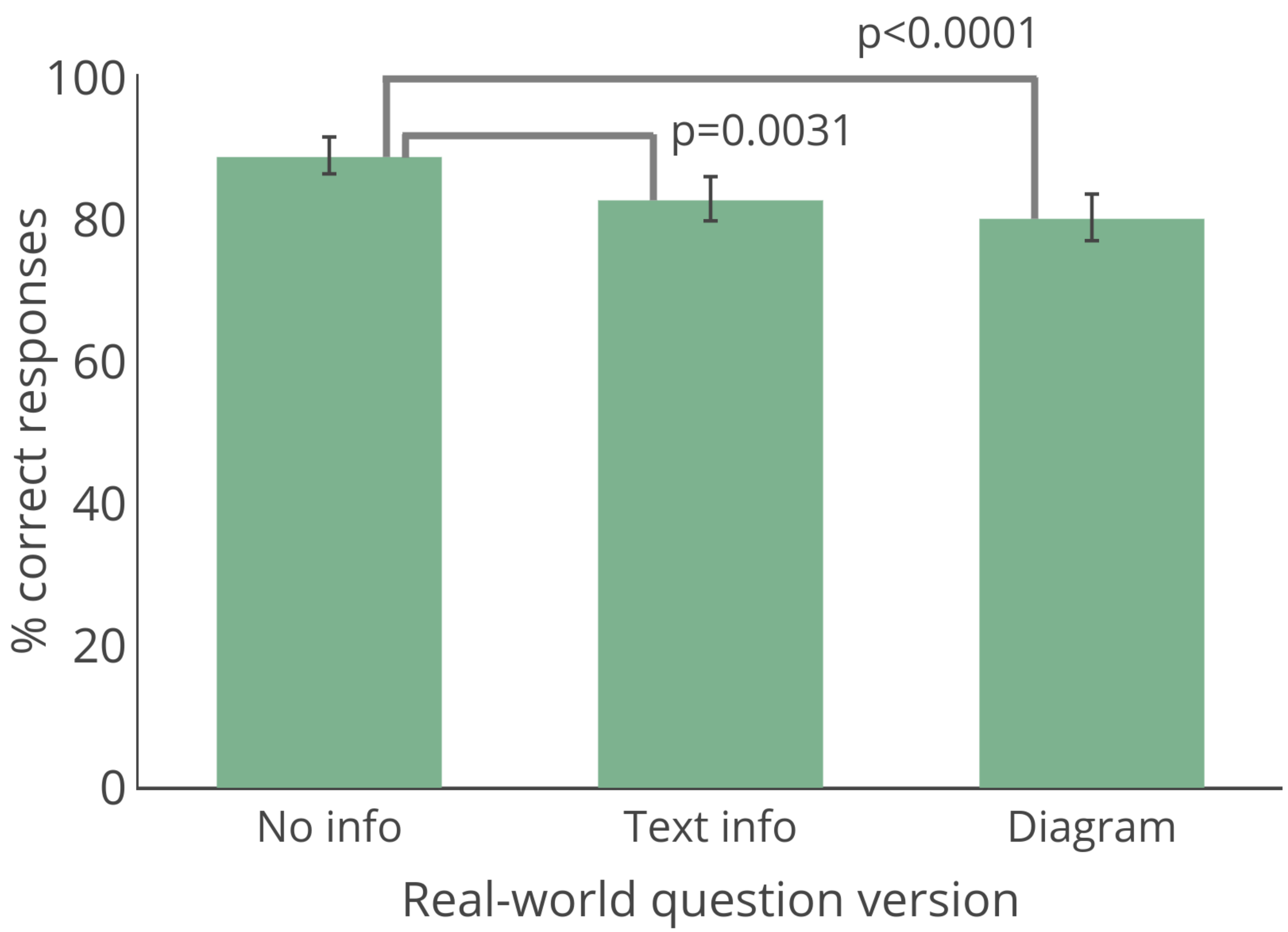
Can causal information aid decision-making in familiar scenarios?

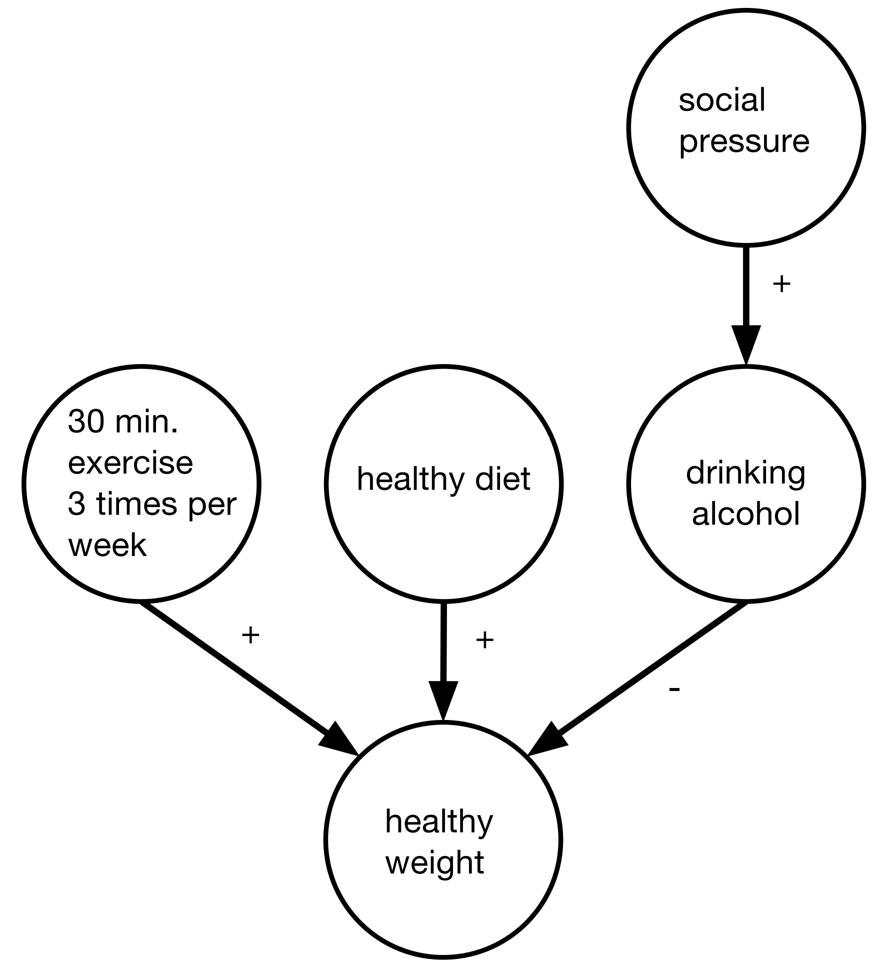
1. Jane just started college and is adjusting to her busy schedule of classes and extracurricular activities. She has heard about the “freshman 15,” where new college students gain 15 pounds during their first year of college. Jane wants to avoid this, while also having fun, making new friends, and leaving time for homework and studying.

What is the ONE thing you think Jane should do to achieve her goal?

- Go for a 30 minute walk every weekend
- Maintain a healthy diet
- Avoid hanging out with friends
- Watch less TV







89% → 80%

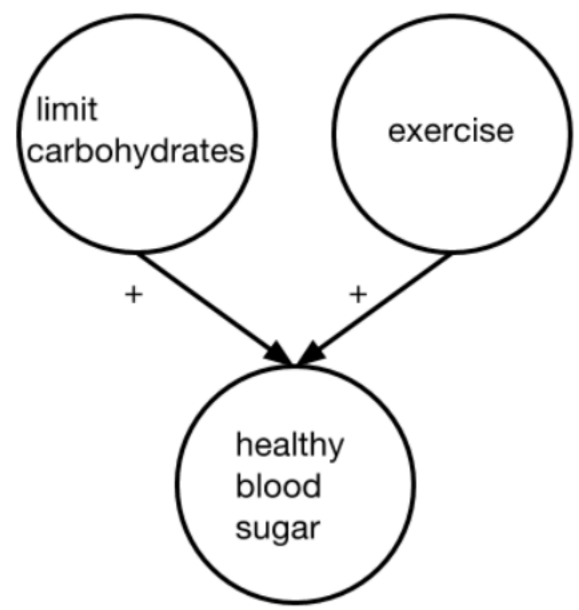
Zheng M, Marsh JK, Nickerson JV, Kleinberg S. How Causal Information Affects Decisions. CRPI (2020)

Are people doing worse
because they have
experience with the domain?

1. Bob was recently diagnosed with Type 2 diabetes. His body does not produce enough insulin, so after a meal, his blood sugar may become dangerously high. Bob does not want to inject insulin, and was relieved when the doctor said his diabetes could be controlled with diet and exercise and ensuring he maintains a healthy weight.

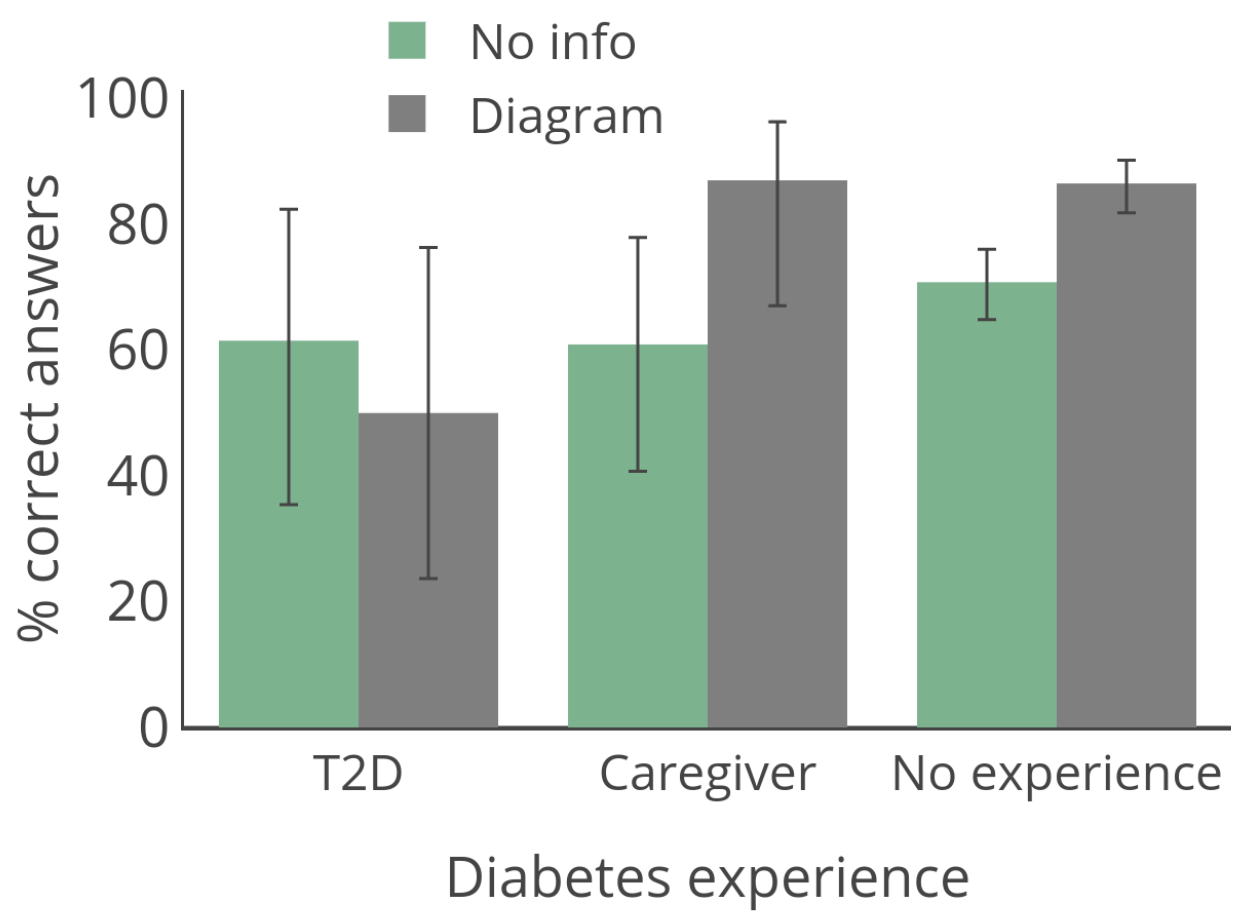
Bob has had a stressful week at work and is looking forward to seeing his friends Friday evening. They usually meet at Bob's favorite fast food restaurant for hamburgers, but now he wonders if that's okay.

What is the BEST suggestion you can give Bob to keep his diabetes under control, and avoid needing insulin injections?



- Walk to dinner
- Order a grilled chicken sandwich instead of a hamburger
- Order a grilled chicken salad and ask his friends to go for a bike ride
- Do what he usually does

Causal information helps the less experienced



Accuracy increases from 71 to 87% for people without diabetes

Are people losing confidence?

Causal information affects confidence in decisions

	No info		Causal model
Experience	83% v. confident 92% correct	→	70% 82% correct
No experience	42% v. confident 87% correct	→	69% 94% correct

What if we change what people think they know?

Kleinberg & Marsh. Tell Me Something I Don't Know. CogSci (2020)

Knowing what you don't know

	No info	Causal model
Subjective	71% correct	71%
Control	75% correct	60%

Knowing what you don't know

	No info	Causal model
Objective	73% correct	72%
Control	72% correct	59%

Action requires causality

But causality alone isn't enough

We need evaluations of utility of algorithms (not just accuracy)

Explainable AI → **Usable** AI

Information must be personalized

Thanks! Teams at @ Columbia, Stevens, Lehigh

Health +AI Lab

Funded by NLM/NIH, NSF, and JSMF

Consciousness

Vital for decision-making, but time consuming to assess and poorly understood

- EEG can help assessment (Claassen J, et al. (2016). Ann Neurol)
- Changes in consciousness associated w/outcomes (Reznik ME, et al. (2018) Neurocrit Care)
- Brain lesions associated w/ impaired consciousness (Rohaut, B. et al. Scientific Report. (in press))

JFK COMA RECOVERY SCALE - REVISED ©2004

Record Form

This form should only be used in association with the "CRS-R ADMINISTRATION AND SCORING GUIDELINES" which provide instructions for standardized administration of the scale.

Patient:	Diagnosis:	Etiology:
Date of Onset:	Date of Admission:	

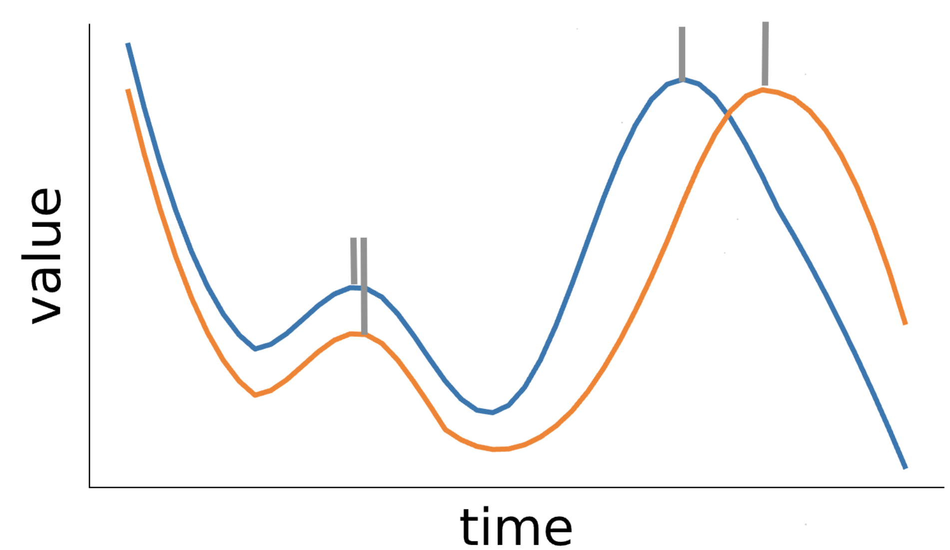
Date																
Week	ADM	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
AUDITORY FUNCTION SCALE																
4 - Consistent Movement to Command *																
3 - Reproducible Movement to Command *																
2 - Localization to Sound																
1 - Auditory Startle																
0 - None																
VISUAL FUNCTION SCALE																
5 - Object Recognition *																
4 - Object Localization: Reaching *																
3 - Visual Pursuit *																
2 - Fixation *																
1 - Visual Startle																
0 - None																
MOTOR FUNCTION SCALE																
6 - Functional Object Use †																
5 - Automatic Motor Response *																
4 - Object Manipulation *																
3 - Localization to Noxious Stimulation *																
2 - Flexion Withdrawal																
1 - Abnormal Posturing																
0 - None/Flaccid																
OROMOTOR/VERBAL FUNCTION SCALE																
3 - Intelligible Verbalization *																
2 - Vocalization/Oral Movement																
1 - Oral Reflexive Movement																
0 - None																
COMMUNICATION SCALE																
2 - Functional: Accurate †																
1 - Non-Functional: Intentional *																
0 - None																
AROUSAL SCALE																
3 - Attention																
2 - Eye Opening w/o Stimulation																
1 - Eye Opening with Stimulation																
0 - Unarousable																
TOTAL SCORE																

BASELINE OBSERVATION AND COMMAND FOLLOWING PROTOCOL ©2004

Commands	Baseline	Trial 1	Trial 2	Trial 3	Trial 4
	1 minute frequency count				
I Object Related Commands					
A. Eye Movement Commands					
Look at the *(object #1)*					
Look at the *(object #2)*					
B. Limb Movement Commands					
Take the *(name object #1)*					
Take the *(name object #2)*					
Kick the *(name object #1)*					
Kick the *(name object #2)*					
II Non-Object Related Commands					
A. Eye Movement Commands					
Look away from me					
Look up *(at ceiling)*					
Look down *(at floor)*					
B. Limb Movement Commands					
Touch my hand					
Touch your nose					
Move your *(object/body part)*					
C. Oral Movement/ Vocalization Commands					
Stick out your tongue					
Open your mouth					
Close your mouth					
Say "ah"					
Spontaneous Eye Opening		Yes:		No:	
Spontaneous Visual Tracking		Yes:		No:	

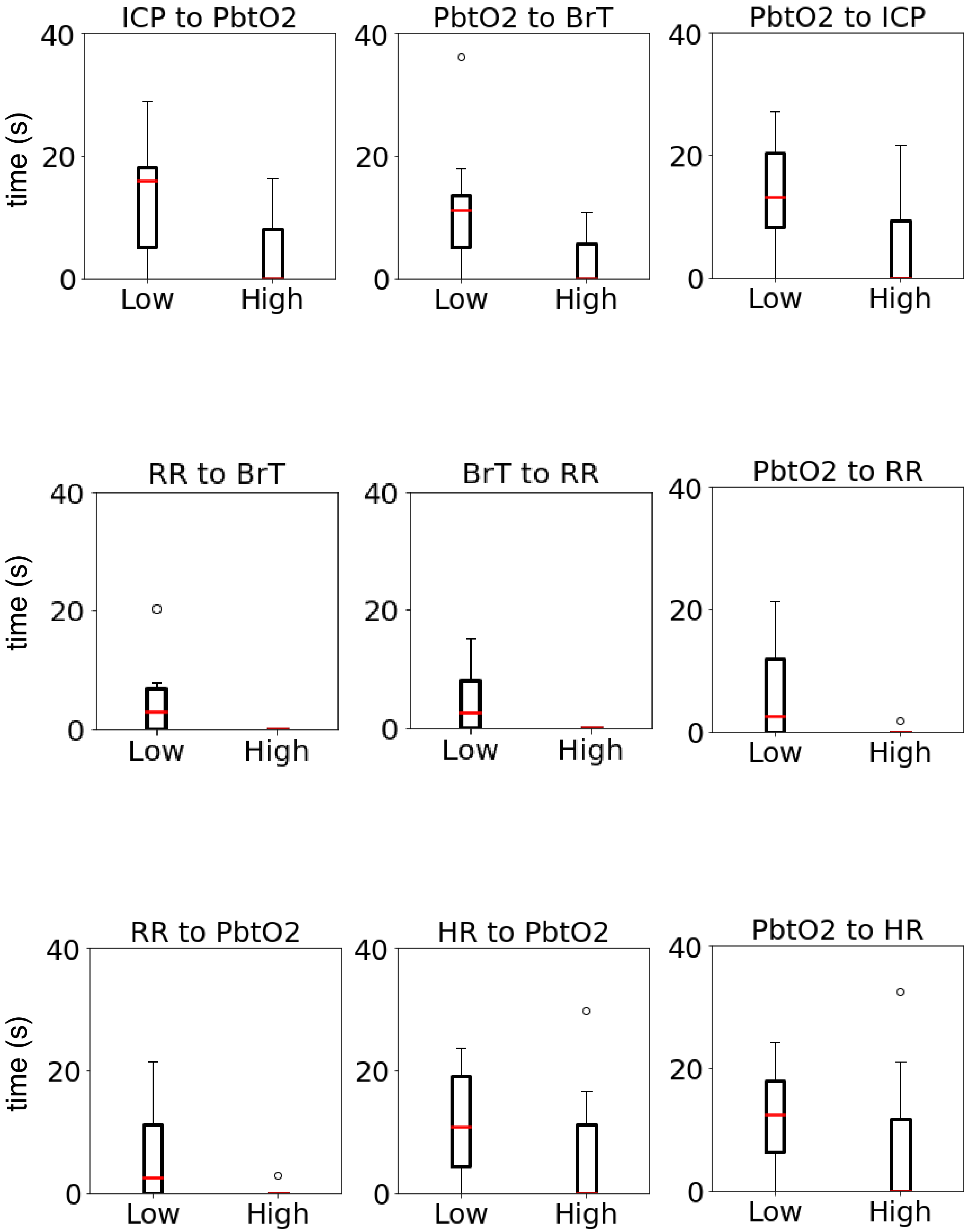
What if we had an automated indicator for consciousness?

Respiratory and cardiovascular signals
Brain-related indicators



Data

- 61 subarachnoid hemorrhage patients in NICU
- 302 assessments of consciousness daily during morning rounds (between 1 and 18 per patient)
- Focusing here on impaired and intact consciousness



BrT: brain temp
PbtO2: brain oxygenation
ICP: intracranial pressure
RR: respiration rate
HR: heart rate

Physiological Variables		p-value	# of Patients		Mean Lag (Seconds)	
From	To	$p < 0.1$	Low	High	Low	High
ICP	PbtO2	0.0117	14	9	13.87	5.06
PbtO2	BrT	0.0329	11	5	11.30	3.31
PbtO2	ICP	0.0343	14	9	13.61	6.65
RR	BrT	0.0360	8	4	5.01	0.00
BrT	RR	0.0360	8	4	4.58	0.00
PbtO2	RR	0.0483	9	6	6.61	0.30
RR	PbtO2	0.0631	9	6	6.25	0.49
HR	PbtO2	0.0666	14	9	11.47	7.00
PbtO2	HR	0.0753	14	9	12.16	8.06

Yavuz, T. Claassen J, and Kleinberg S. Lagged Correlations among Physiological Variables as Indicators of Consciousness in Stroke Patients. (under review).