Implementing machine learning based solutions into real-life: Everything you need to know in 29 minutes

Mike Fralick, MD, PhD, SM, FRCPC

General Internist, Mount Sinai Hospital, University of Toronto

Assistant Professor, Department of Medicine, University of Toronto

@Fralickmike

Machine Learning Course, University of Toronto

20 Oct 2021

UNIVERSITY OF TORONTO

Mount Sinai Hospital

Disclosures

- Conflicts of interest
    - ProofDx
    - Honouraria received from NEJM and Lancet
    - Board member for NEJM Evidence

St. Michael's
Inspired Care.
Inspiring Science.

UNIVERSITY OF TORONTO
Eliot Phillipson Clinician Scientist Training Program

DSECT
Drug Safety & Effectiveness Cross-Disciplinary Training

CIHR IRSC

ROYAL COLLEGE
OF PHYSICIANS AND SURGEONS OF CANADA

ODPRN
ONTARIO DRUG POLICY RESEARCH NETWORK

Objectives

- Provide a foundation on study design [epidemiology]
- Provide examples of implementing ML studies into clinical care at various hospitals in Ontario

Crash-course in epidemiology

Most important step: coming up with a great Research Question

- P
- I
- C
- O

Research Question

- Population
- Intervention [for RCT] Exposure [for non-RCT]
- Comparator or Control group
- Outcome

What type of question are you asking?

- Causal question: randomized trial
- “Causal” question: cohort or case control

What type of question are you asking?

- Causal question: randomized trial
- Association question: cohort or case control

What type of question are you asking?

- Causal question: randomized trial
- Association question: cohort or case control
- Prediction question: cohort + fancy stats/ML

Now let's talk about study design

Study designs

- “Experimental”
  - Randomized controlled trials

The magic of randomization

- Randomization allows for:
  - Removal of selection bias
  - Balancing of measured confounders
  - Balancing of unmeasured confounders
  - Everyone has the same time-zero

ORIGINAL ARTICLE

Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes

Ur
Bl
Pr
PF

Characteristic*	Placebo (N = 2333)	Empagliflozin 10 mg (N = 2345)
Age – years	63.2 ± 8.8	63.0 ± 8.6)
Male – no. (%)	1680 (72.0)	1653 (70.5)
Race – no. (%)		
White	1678 (71.9)	1707 (72.8)
Asian	511 (21.9)	505 (21.5)
Black/African-American	120 (5.1)	119 (5.1)
Other/Missing	24 (1.0)	14 (0.6)
Ethnicity – no. (%)		
Not Hispanic or Latino	1912 (82.0)	1909 (81.4)
Hispanic or Latino	418 (17.9)	432 (18.4)
Missing	3 (0.1)	4 (0.2)
Region – no. (%)		
Europe	959 (41.1)	966 (41.2)
North America (plus Australia and New Zealand)	462 (19.8)	466 (19.9)
Asia	450 (19.3)	447 (19.1)
Latin America	360 (15.4)	359 (15.3)
Africa	102 (4.4)	107 (4.6)
Weight – kg	86.6 ± 19.1	85.9 ± 18.8
Body mass index – kg/m^2†	30.7 ± 5.2	30.6 ± 5.2
CV risk factor – no. (%)	2307 (98.9)	2333 (99.5)
Coronary artery disease	1763 (75.6)	1782 (76.0)

Zinman B et al., *NEJM* 2015

ORIGINAL ARTICLE

Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes

Unmeasured variable	Placebo	Empagliflozin
Blue eyes	10%	10%
Prior DKA	2%	2%
PRSS1 Gene	1%	1%

Zinman B et al., *NEJM* 2015

Study types

- Experimental
  - Randomized controlled trials

- Observational
  - Ecological study
  - Cross-sectional study
  - Cohort study
  - Case-control study
  - Case-crossover

Clinical Epidemiology: Everything you need to know in 59 Minutes

Mike Fralick, MD, PhD, MSc

@Fralickmike

UNIVERSITY OF TORONTO

Mount Sinai Hospital

Study types

- Experimental
  - Randomized controlled trials

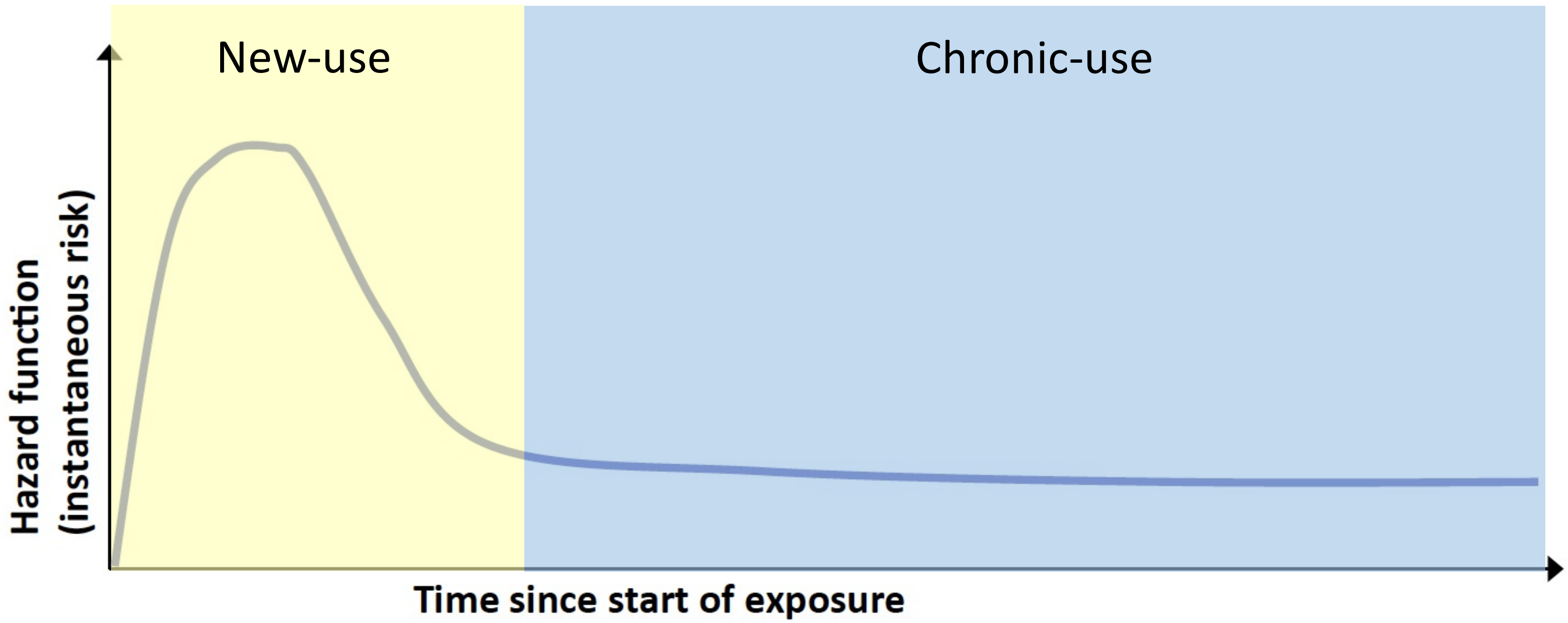
- Observational
  - Ecological study
  - Cross-sectional study
  - **Cohort study**
  - Case-control study
  - Case-crossover

Cohort studies

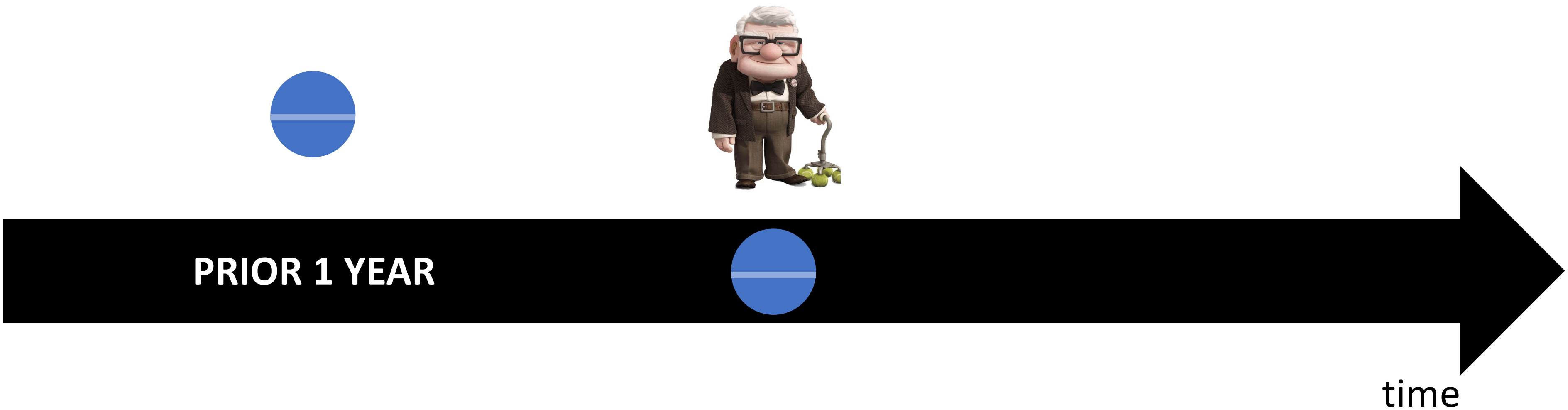
- Defined by a cohort entry event and people are followed over time
- Cohort of medical students
- Cohort of people at the talk today

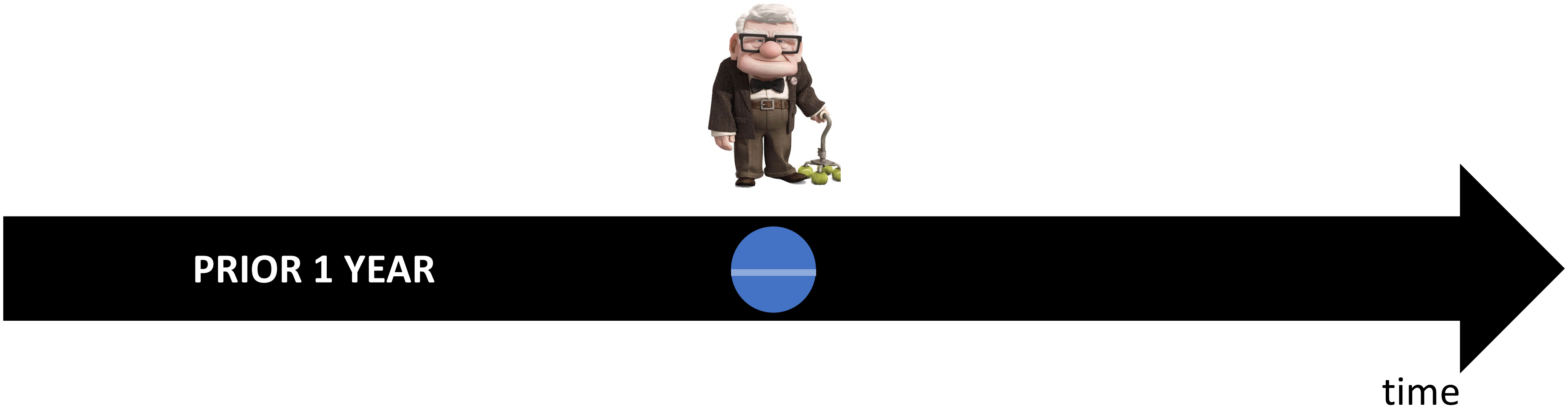

time

Time-varying hazards




PRIOR 1 YEAR
time


PRIOR 1 YEAR
time


PRIOR 1 YEAR
time


PRIOR 1 YEAR
time

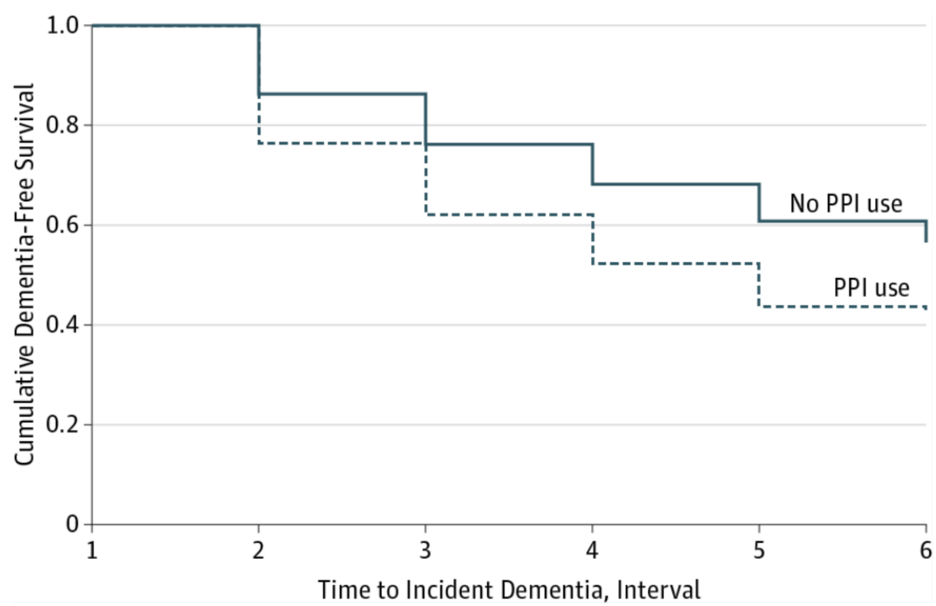
Original Investigation

FREE

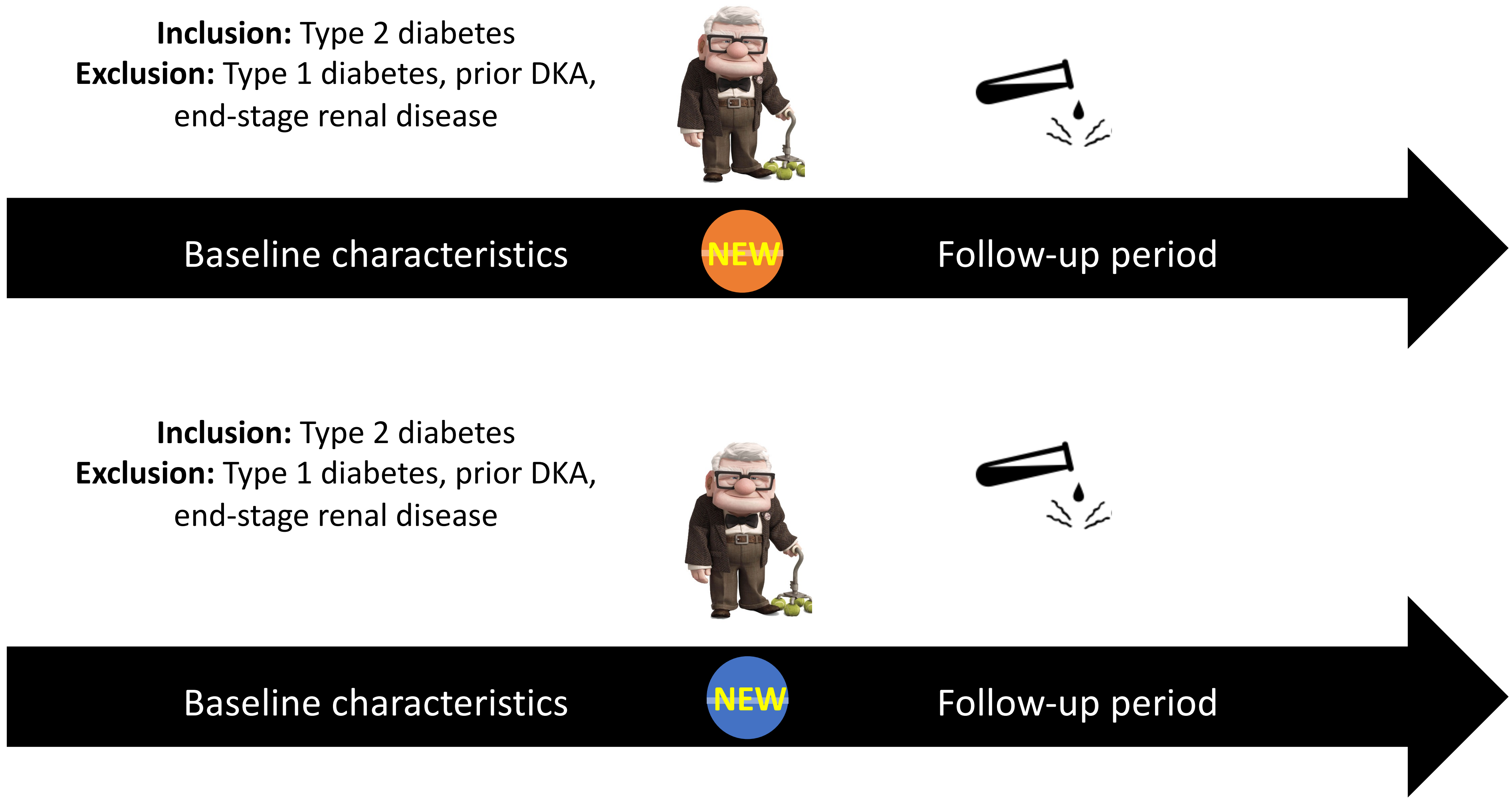
April 2016

Association of Proton Pump Inhibitors With Risk of Dementia

A Pharmacoepidemiological Claims Data Analysis



Higher risk of dementia with PPI


Inclusion: Type 2 diabetes
Exclusion: Type 1 diabetes, prior DKA, end-stage renal disease
Baseline characteristics
NEW
Follow-up period
Inclusion: Type 2 diabetes
Exclusion: Type 1 diabetes, prior DKA, end-stage renal disease
Baseline characteristics
NEW
Follow-up period

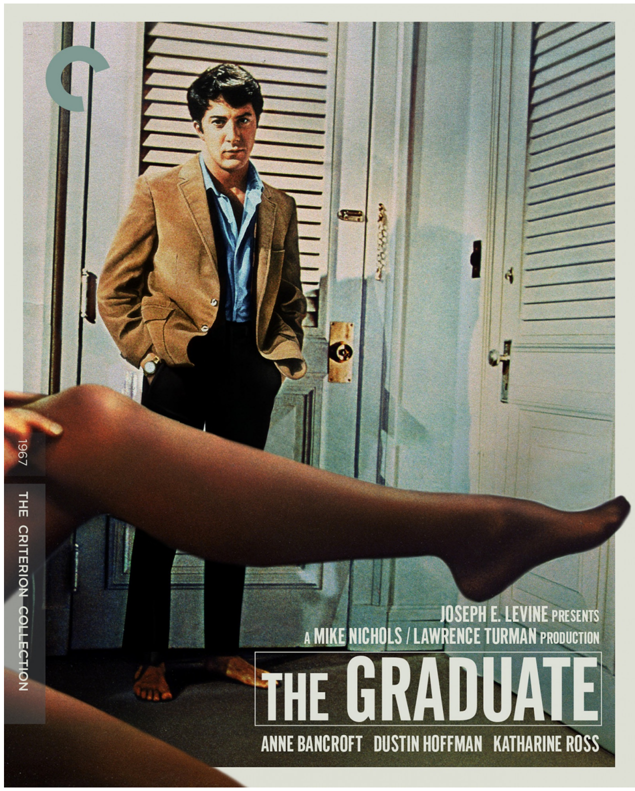
But observational studies don’t randomize participants so how can we prevent bias?

Confounding control by design

- **N**ew user design
- **A**ctive comparator
- **R**elevant confounders identified
- **C**onfounders adjusted for
- **O**utcome identifiable and valid
- **S**ensitivity analyses demonstrate robustness

Ray WA, Am J Epidemiol, 2003
Schneeweiss et al, JAMA 2018
Fralick et al., JAMA IM, 2019

Preventing bias, beyond study design

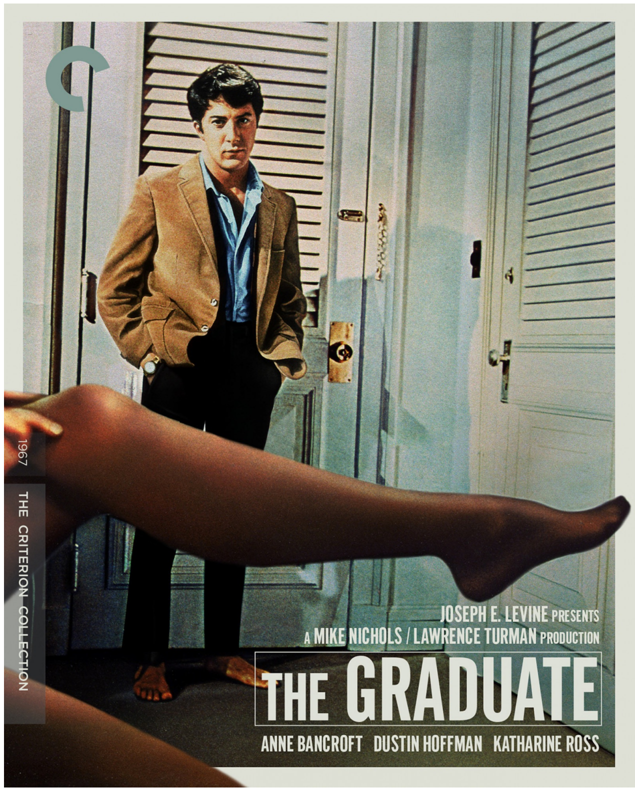


- **MRS. R**obinson

- **M**atching
- **R**estriction
- **S**tratification

- **R**egression

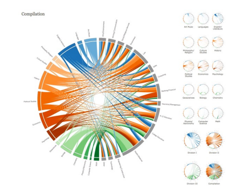
Dear MRS. Robinson



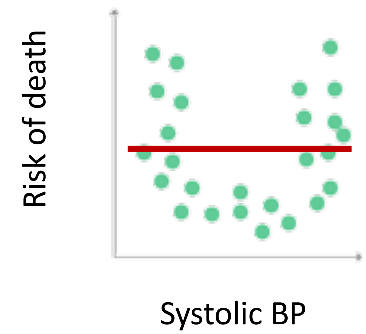
- **De**sign
- **M**atching
- **R**estriction
- **S**tratification
- **R**egression

Limitations of regression

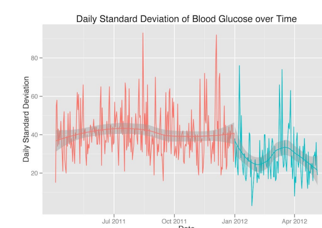
- Over-fitting with high-dimensional data
  - “10 to 1” rule



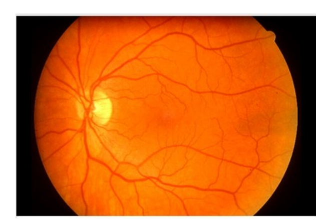
- Handling of non-linear relationships



- Handling of time-varying variables



- Inability to interpret images



Machine Learning

James G , et al. An introduction to statistical learning

- Definition: a form of artificial intelligence which mines data for patterns. These patterns can provide a rich understanding of the data and potentially aid in clinical prediction

- Supervised Learning

- Unsupervised Learning

Machine Learning

Supervised Learning

Can an automated algorithm detect diabetic retinopathy from retinal photographs ?

Research

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

DESIGN AND SETTING A specific type of neural network optimized for image classification called a deep convolutional neural network was trained using a retrospective development data set of 128 175 retinal images, which were graded 3 to 7 times for diabetic retinopathy, diabetic macular edema, and image gradability by a panel of 54 US licensed ophthalmologists and ophthalmology senior residents between May and December 2015. The resultant algorithm was validated in January and February 2016 using 2 separate data sets, both graded by at least 7 US board-certified ophthalmologists with high intragrader consistency.

EXPOSURE Deep learning–trained algorithm.

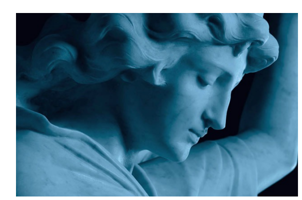
Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables

Emma Ahlqvist, Petter Storm, Annemari Käräjämäki, Mats Martinell*, Mozhgan Dorkhan, Annelie Carlsson, Petter Vikman, Rashmi B Prasad, Dina Mansour Aly, Peter Almgren, Ylva Wessman, Nael Shaat, Peter Spégel, Hindrik Mulder, Eero Lindholm, Olle Melander, Ola Hansson, Ulf Malmqvist, Åke Lernmark, Kaj Lahti, Tom Forsén, Tiinamaija Tuomi, Anders H Rosengren, Leif Groop*

Methods We did data-driven cluster analysis (k-means and hierarchical clustering) in patients with newly diagnosed diabetes (n=8980) from the Swedish All New Diabetics in Scania cohort. Clusters were based on six variables (glutamate decarboxylase antibodies, age at diagnosis, BMI, HbA_{1c}, and homoeostatic model assessment 2 estimates of β-cell function and insulin resistance), and were related to prospective data from patient records on development

Risk of Unintentional Severe Hypoglycemia in Hospital (RUSHH)

St. Michael's
Inspired Care.
Inspiring Science.
LKS-CHART



Mr. B

ID: 80M admitted with pneumonia.

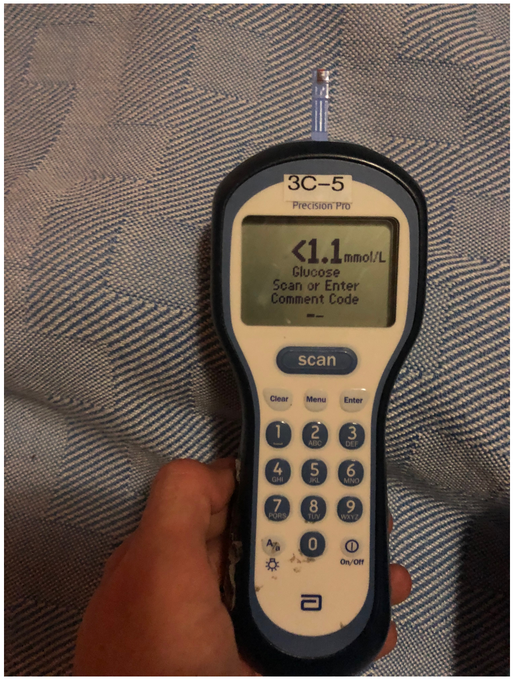
Medical history: diabetes, coronary artery disease, dialysis

Medications: insulin, aspirin, atorvastatin, metoprolol (new), moxifloxacin (new)

By day 5 he recovered from his pneumonia and was planned for discharge the following day (Friday).

Friday at 8AM we got a STAT page that he was unresponsive.

We assessed him, a bedside blood glucose was performed.



Hypoglycemia Risk Score

Predicts 12-month risk of hypoglycemic episodes in T2DM patients.

Pearls/Pitfalls | Why Use

Intermediate risk

1-5% 12-month risk of hypoglycemia admission

Copy Results | Next Steps

Severe or end-stage kidney disease eGFR ≤29 by CKD-EPI Creatinine	No	**Yes**
Age	<77 years	**≥77 years**

Please fill out required fields.

DRUGS

Orders

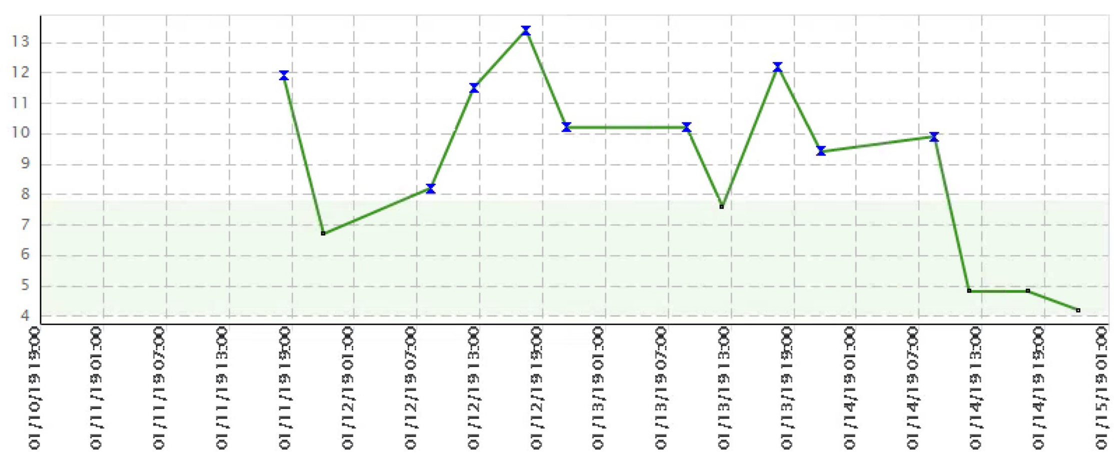
MD Notes

Nursing Notes

LABS

ASSESS PLAN
Senior Resident
Impression/Plan
Overall, 56YF with ESRD on IHD, bipolar disorder, HTN, DLD, DM2 on insulin presenting with confusion and decreased LOC, acute tonic clonic seizure after having missed 2 episodes of dialysis. Her investigations are notable for: elevated urea, hyperkalemia with no ECG changes, Cr elevated (although on dialysis), and mild troponitis. It is difficult to decipher the timeline of events- either she was confused and then missed her meds/ dialysis and underwent consequential seizures, or, she seized and in her post ictal state missed dialysis which lead to further seizures. The plan moving forward is as follows:

Glucose POC (Lifescan/Abbott/Nova) mmol/L

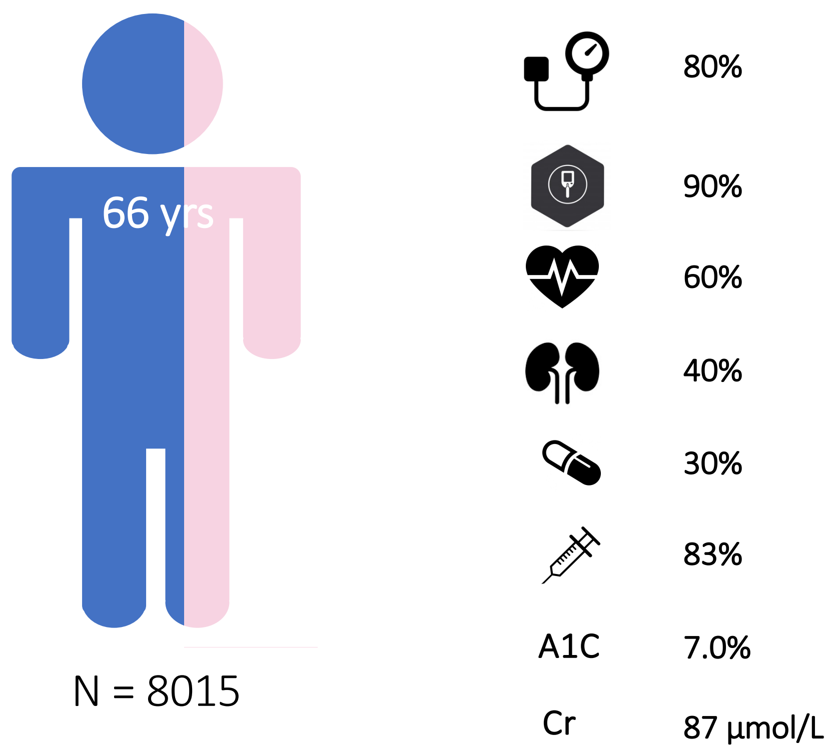


01/15/19
12:56
01151776
1.7 LL

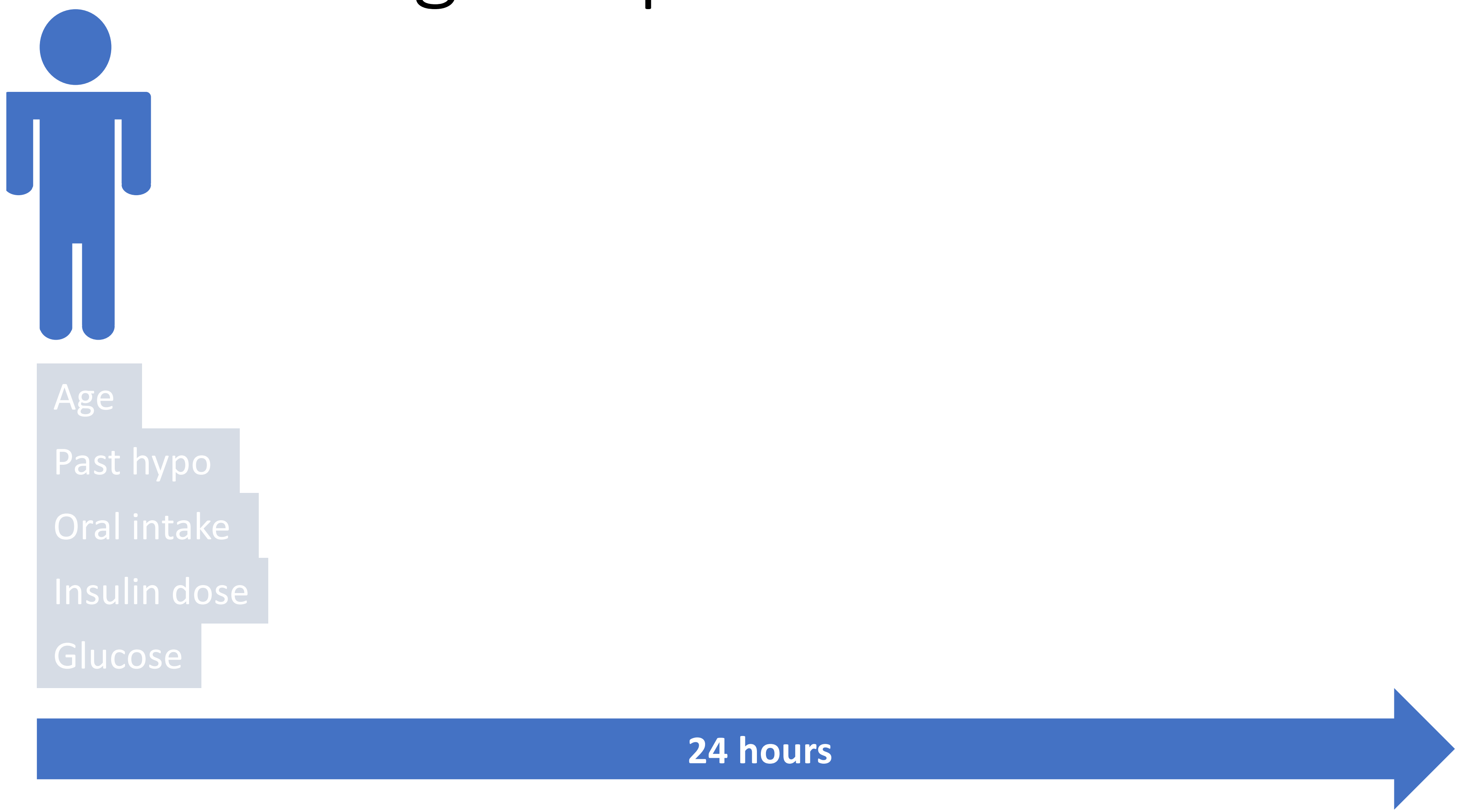
Model building

- **Training:** 2013 – 2017
- **Validating:** 2017 – 2018
- **Testing:** 2019 – 2019
- **Implementation:** 2020

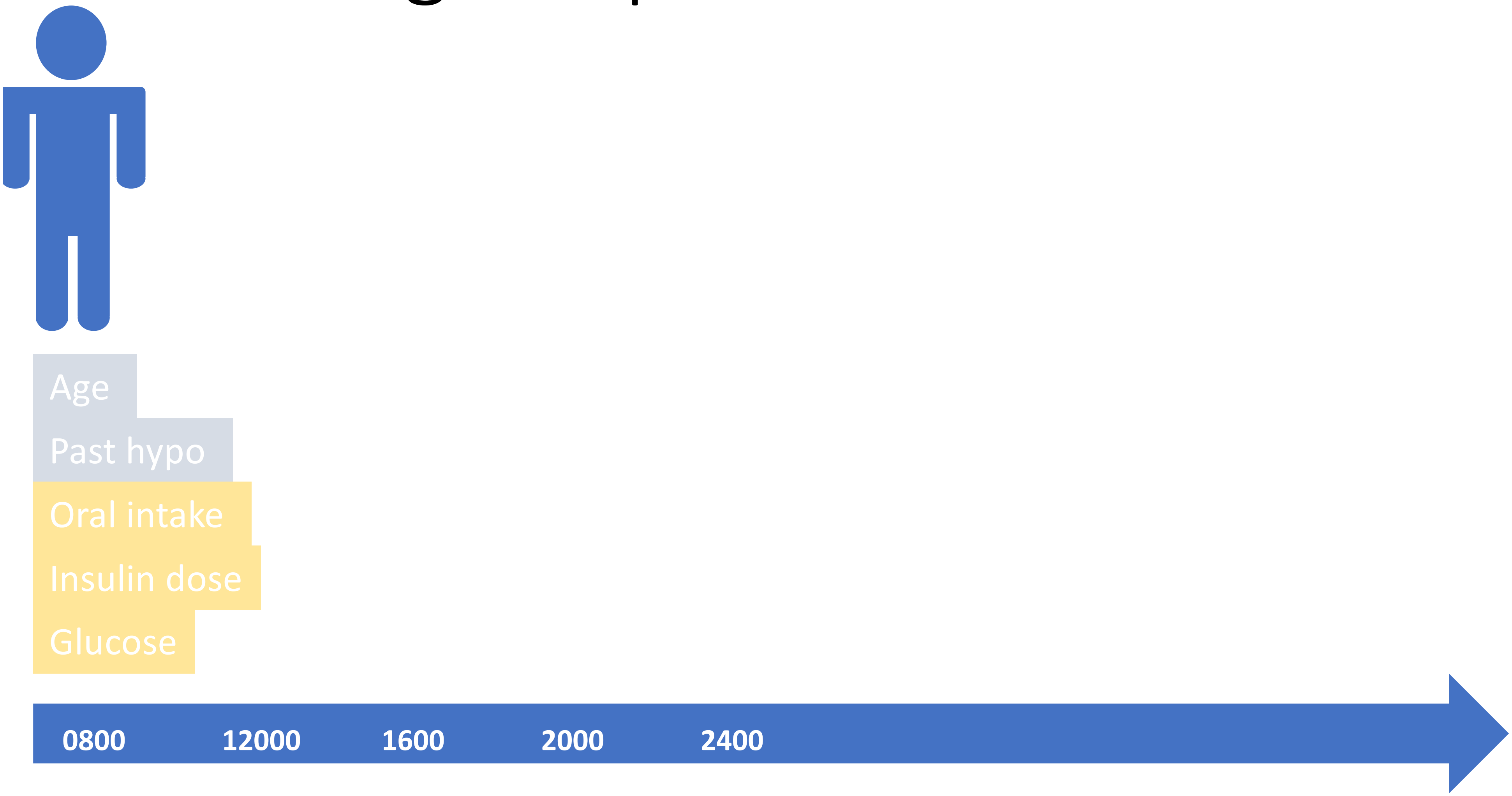
Patient characteristics



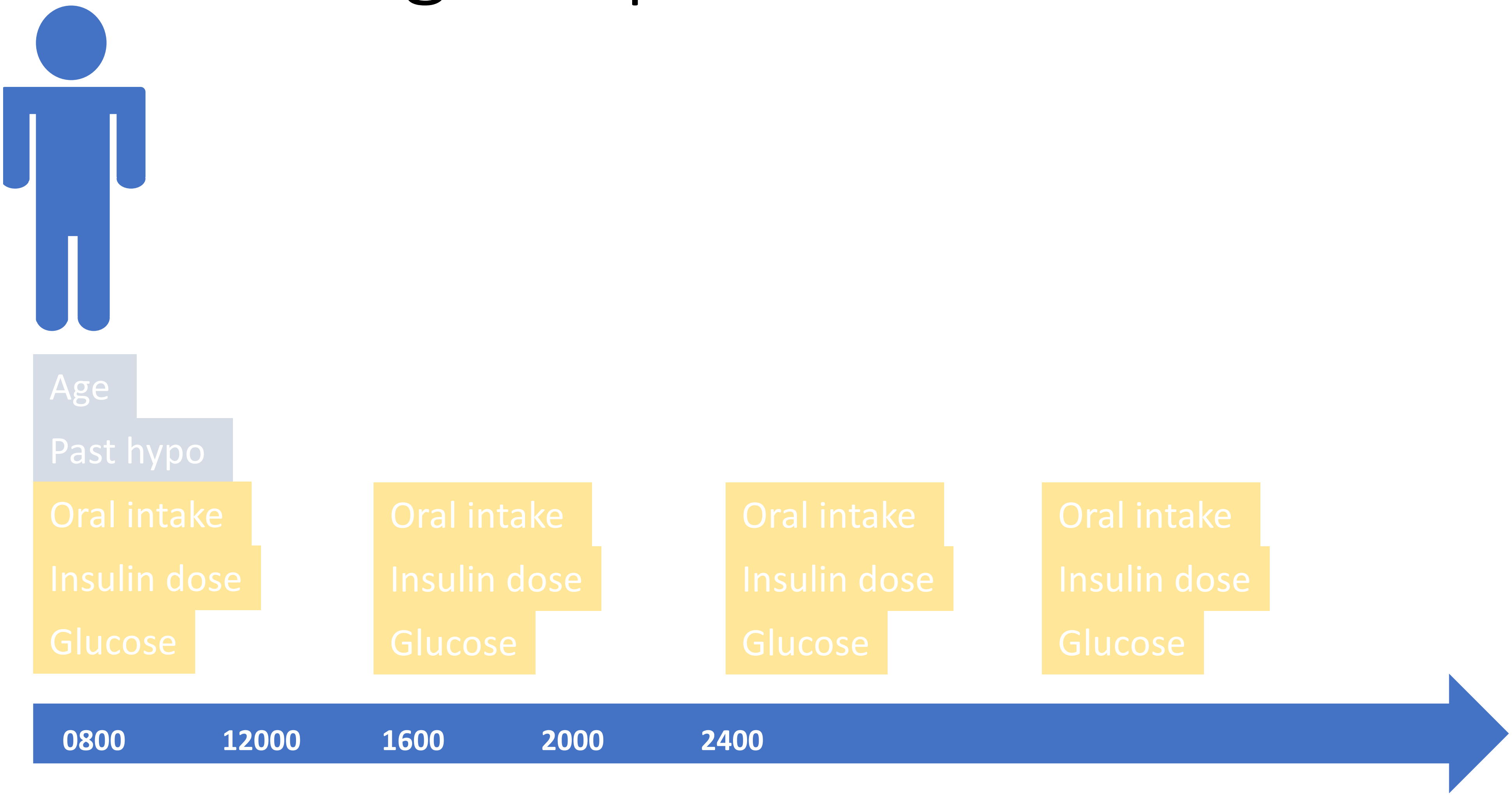
Visualizing the problem



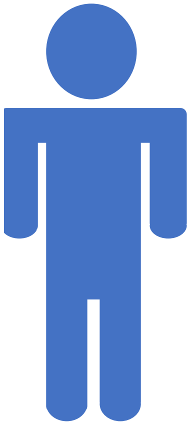
Visualizing the problem



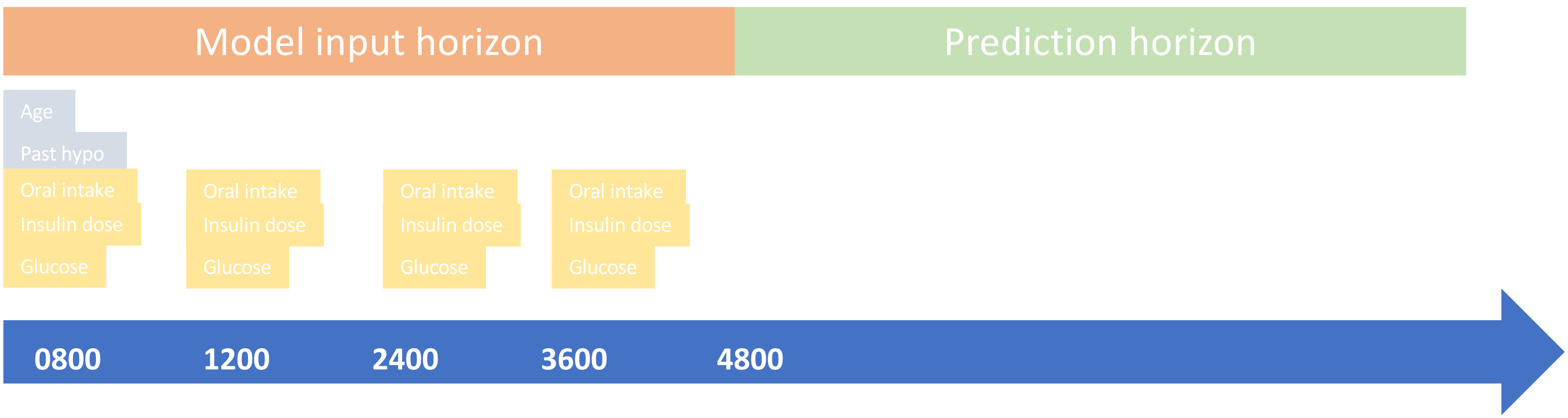
Visualizing the problem



Visualizing the problem



Interpretation: model provides a prediction for the subsequent 24 hours based on the preceding 24 hours (or more) of data



How our Model Works

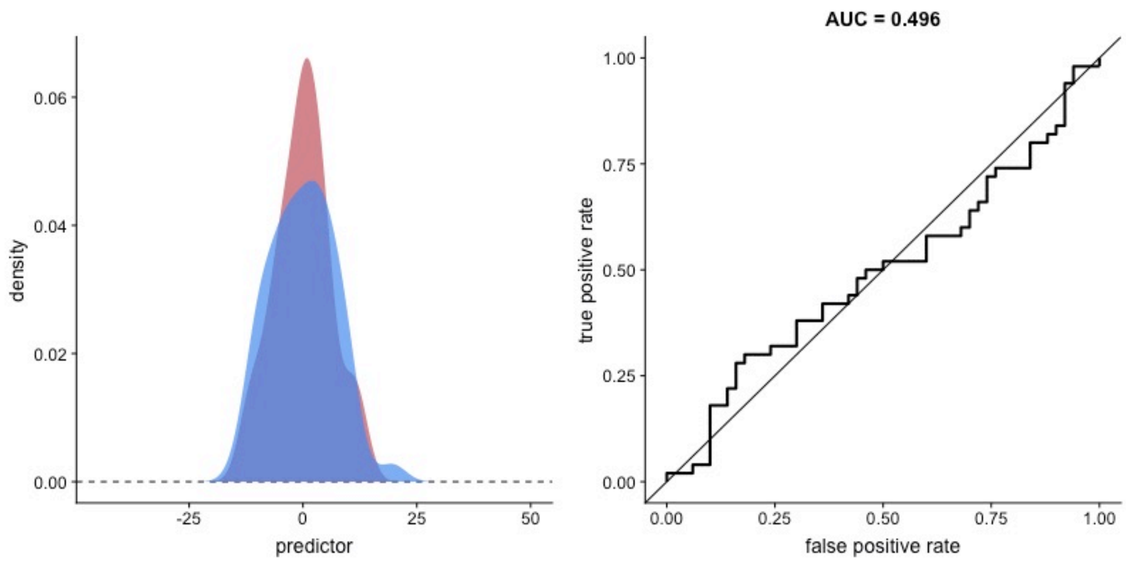
- **Prediction time:** 08:00
- **Time horizon:** 24 hours
- **Analytic approach:**
  - Penalized regression
  - Gradient boosted trees
  - Recurrent neural network
  - Ensemble

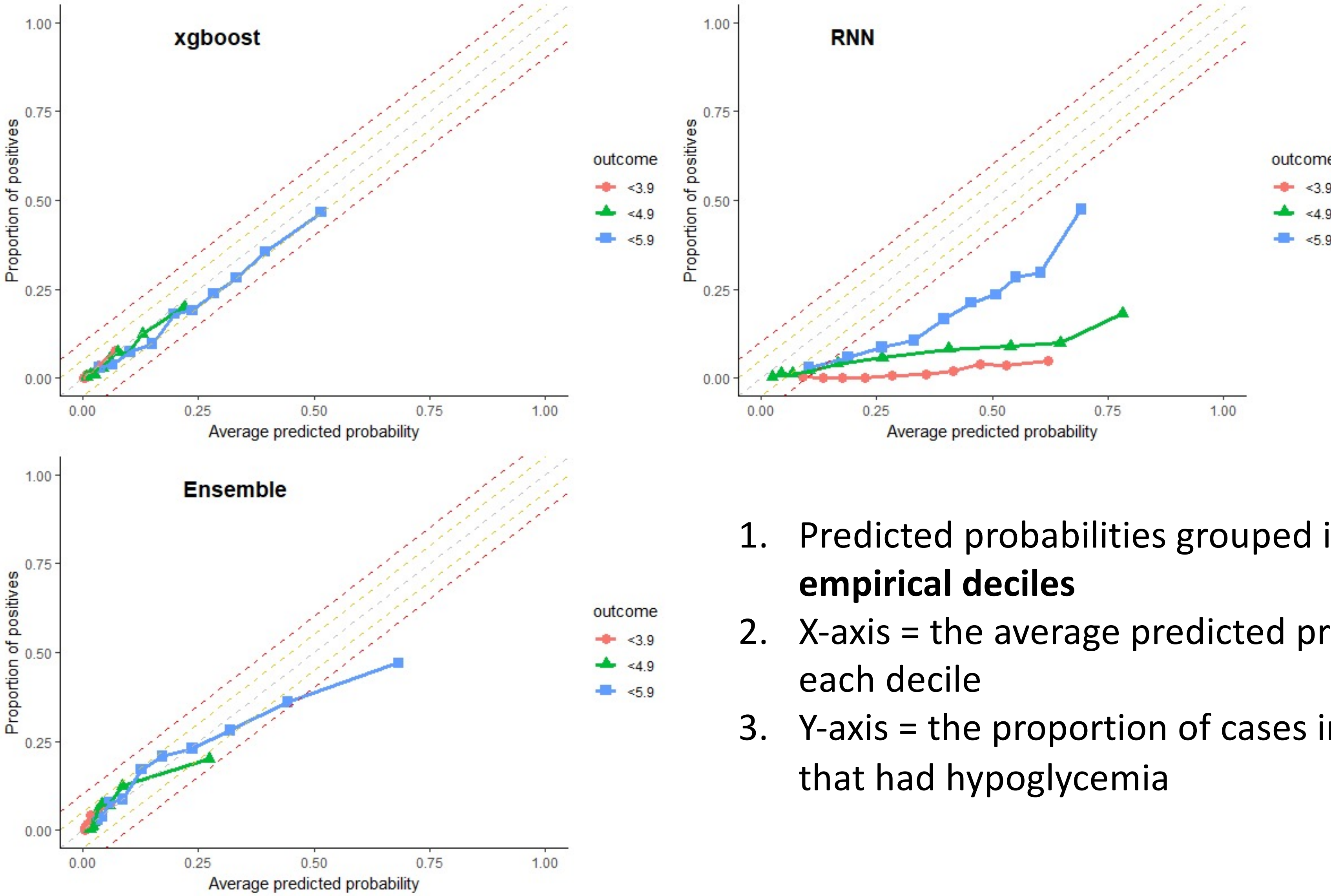
How the Model Works

- *Time-varying features*:
  - **Glucose** (glucose_value, glucose_lowest, glucose_highest)
  - **Measures** (measure_temperature, measure_sbp, …)
  - **Insulin** (insulin_short, insulin_long, insulin_combo)
  - **Oral intake** (oral_intake_pct, oral_intake_is_vomit, …)
  - **NPO** (npo_mention_nursing_note, npo_midnight_nursing_note, …)
  - **Medications** (med_hypoglycemia, med_hypoglycemia_high_risk, …)
- *Static features*:
  - **Demographics** (age, gender)
  - **Dialysis** (dialysis_acute, dialysis_chronic)
  - **Baseline measures** (baseline_creatinine, baseline_albumin, …)
  - **Patient history** (history_hypoglycemia, history_diabetes, …)

Model AUC

Model	Train	Validation	Test
xgboost 3.9	0.93 (0.5)	0.81 (0.08)	0.83 (0.06)
RNN 3.9	0.78 (0.07)	0.74 (0.05)	0.78 (0.05)
Ensemble 3.9	0.92 (0.49)	0.8 (0.08)	0.83 (0.06)


density
0.06
0.04
0.02
0.00
-25
0
25
50
predictor
AUC = 0.496
true positive rate
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
false positive rate



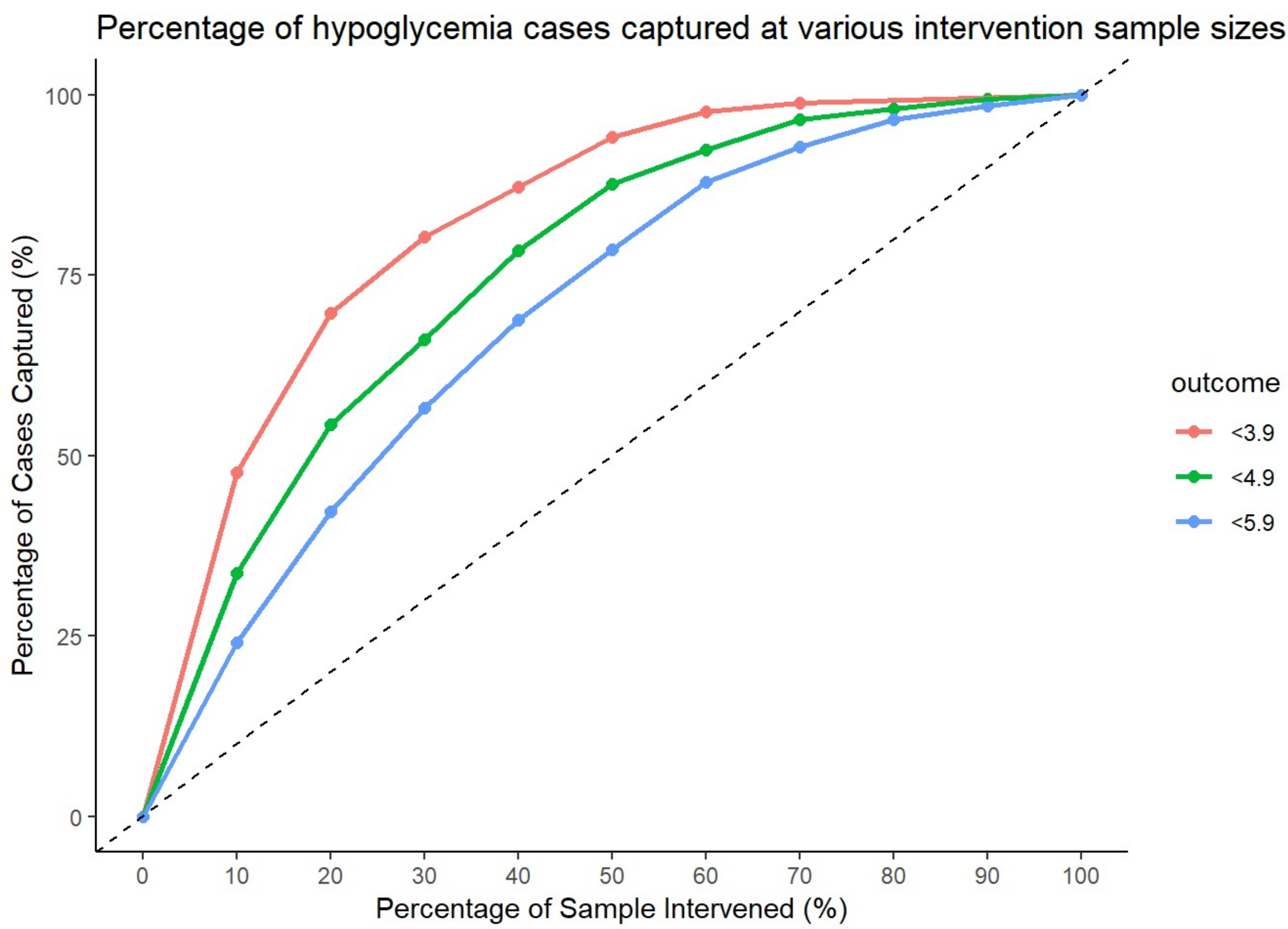
1. Predicted probabilities grouped into **empirical deciles**
2. X-axis = the average predicted probability in each decile
3. Y-axis = the proportion of cases in each group that had hypoglycemia

Other model metric

TABLE 3 medicine and cardiovascu

Glucose thre (mmol/L)	GIM LASSO	XGBo	XGBoost	RNN
2.9	0.023	0.023	.018	0.0252
	0.036	0.029	.019	0.0289
	0.041	0.048	.021	0.0306

Positiv cks !!


Percentage of hypoglycemia cases captured at various intervention sample sizes
Percentage of Cases Captured (%)
100
75
50
25
0
0
10
20
30
40
50
60
70
80
90
100
Percentage of Sample Intervened (%)
outcome
<3.9
<4.9
<5.9

Model implementation

- **Iterative process that requires constant communication with your end user**

Model implementation pearls

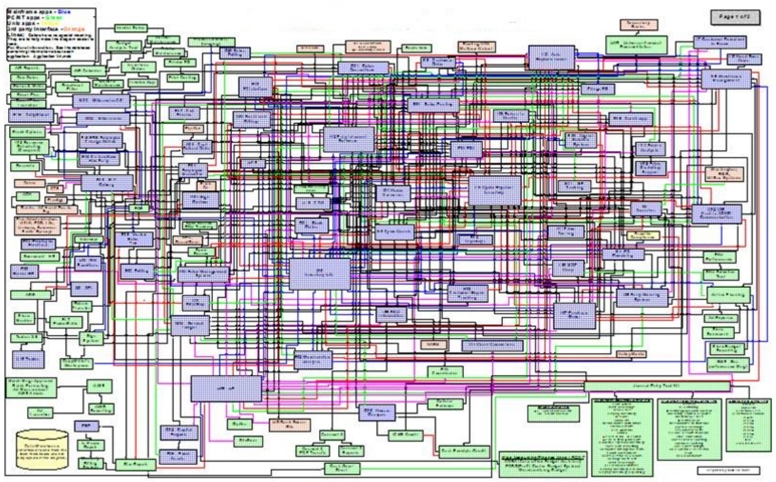
1. **It is EASY to run a model on cleaned retrospective data.**

```r
library(gbm)
# train GBM model for PO3
set.seed(123)
gbm.fit <- gbm(
  formula = PO3 ~ .,
  distribution = "bernoulli",
  data = DRAWS_GBTprepPO3,
  n.trees = 3010,
  interaction.depth = 3,
  shrinkage = 0.001,
  cv.folds = 10
)
summary(gbm.fit, order=TRUE, las=1)
```

Model implementation pearls

1. **It is EASY to run a model on cleaned retrospective data.**
2. **Implementing a model requires a completely different skillset**

Implementing a model



Model implementation pearls

1. **It is EASY to run a model on cleaned retrospective data.**
2. **It is HARD to implement a model and ensure your data pipelines are in place and working**
3. **Spend time with your end-user in their native environment**

Model implementation pearls

1. **It is EASY to run a model on cleaned retrospective data.**
2. **It is HARD to implement a model and ensure your data pipelines are in place and working**
3. **Spend time with your end-user in their native environment**
4. **Understand their daily workflow and what "pisses them off"**

Model implementation pearls

1. **It is EASY to run a model on cleaned retrospective data.**
2. **It is HARD to implement a model and ensure your data pipelines are in place and working**
3. **Spend time with your end-user in their native environment**
4. **Understand their daily workflow and what "pisses them off"**
5. **See what their data looks like from the "front end", can you find it in the "back-end"?**

RUSHH implementation

- Daily email with list of patients at highest decile of risk of severe hypoglycemia

David Dai
Sr. Data Scientist

Kasthuri Karunanithi
Research Assistant I

Sebnem Kuzulugil
Director, Data Integration and Governance

Muhammad Mamdani
Vice-President

Neil Mistry
Sr. Data Scientist

Joshua Murray
Director, Advanced Analytics

Chloé Pou Prom
Sr. Data Scientist

Colin Purcell
System Administrator

Predicting ER Volumes

- **Research question:** Can we leverage machine learning techniques to predict how many patients will show up to the ER each day?

- **Study design:** Prospective cohort study at 3 hospitals in the greater Toronto area

- **Variables of interest:** historical data, holidays, weather, major events in Toronto

Analytic approach

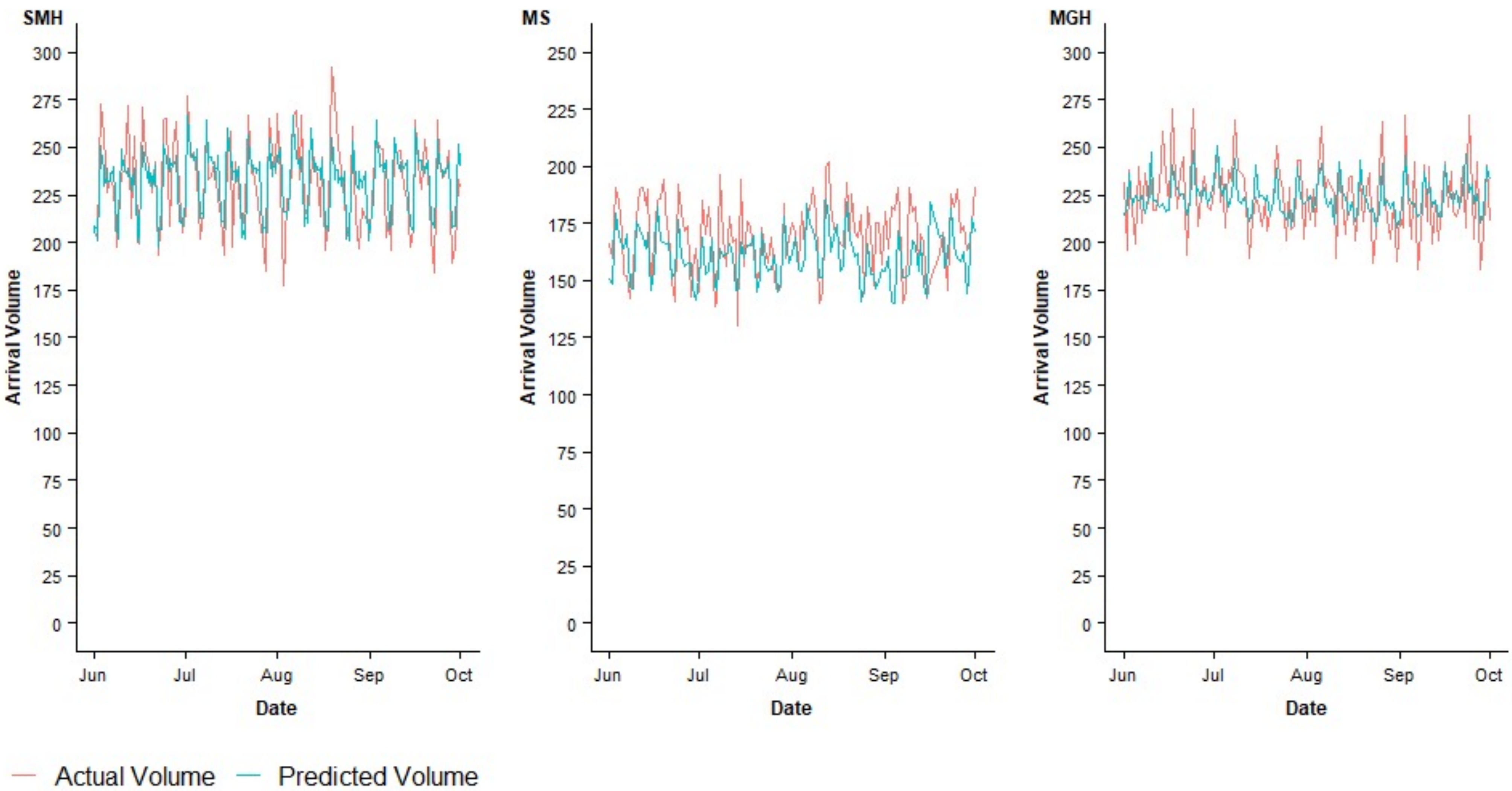
- Neural network
- Random Forest
- ARIMA model
- Exponential smoothing state space model

- Models were trained and validated using data from 2016 to 2019 and the results provided are from 2019 – 2020 [test set]

Exponential smooth state model

$$\hat{y}_{t+h|t} = \ell_t + s_{t+h-m(k+1)}$$
$$\ell_t = \alpha\left(y_t - s_{t-m}\right) + (1-\alpha)\left(\ell_{t-1}\right)$$
$$s_t = \gamma\left(y_t - \ell_{t-1}\right) + (1-\gamma)s_{t-m}$$

1. h is the horizon of forecast; in our model, h is 7
2. k is the integer part of (h−1)/m
3. ℓ_t is the level equation which represents a weighted average between the seasonally adjusted observation $y_t - s_{t-m}$. The formula is derived from the weighted average equation: $\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$
4. s_t is the seasonal equation which represents a weighted average between the current seasonal index, $(y_t - \ell_{t-1})$, and the seasonal index of the same season last year (i.e., m time periods ago)


SMH
MS
MGH
Arrival Volume
Date
Jun
Jul
Aug
Sep
Oct
Actual Volume
Predicted Volume

Main findings

- About 95% accurate at predicting ER volumes
- Also able to predict
  - Level of patient acuity
  - Number of mental health related visits


Sautle Ste Marie

Implementation in the Sault

- End-users: nurse managers, ER doctors
- Plumbers: IT, data engineers
- Analysts: David Dai, Yang Zhu

Input from end-users

- **Nursing manager**
  - Nurses call in sick every day and we need to decide, do we replace the sick call?
  - We need something that is accurate, reliable, and simple

- **Emergency medicine doctors**
  - It would be great if we can predict far in advance so that we can schedule accordingly
  - We also have an ER doctor on back-up each day, but we have now reliable or pre-emptive method of knowing when to call them in

edForecastingApp

ED Arrivals

2021-10-20

138 — 97 / 41

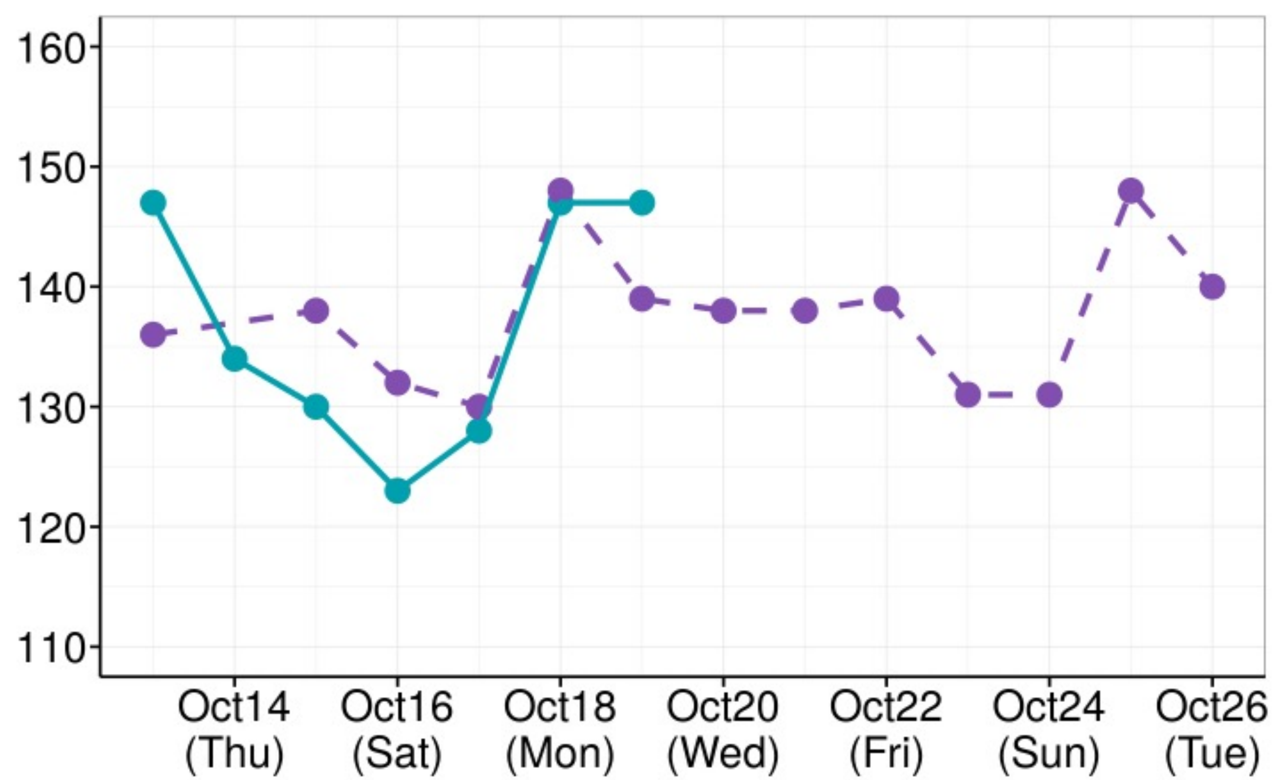
2021-10-21

138 — 98 / 40

2021-10-22

139 — 98 / 41

Daily Arrivals: Forecasted and Actual



Daily arrivals for September 2021

Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1 **144** (130)	2 **132** (133)	3 **149** (132)	4 **125** (129)
5 **118** (127)	6 **147** (144)	7 **149** (135)	8 **121** (134)	9 **118** (133)	10 **152** (130)	11 **128** (128)
12 **126** (128)	13 **131** (144)	14 **134** (133)	15 **130** (130)	16 **143** (130)	17 **144** (134)	18 **115** (129)
19 **109** (126)	20 **139** (140)	21 **143** (132)	22 **136** (132)	23 **132** (135)	24 **126** (135)	25 **125** (126)
26 **143** (124)	27 **165** (145)	28 **139** (141)	29 **131** (137)	30 **144** (134)		

Calendar cell values represent: **Actual arrivals** (Forecasted arrivals)
Calendar cell colours represent: absolute forecasted error of <5 arrivals, and >=20 arrivals

Pearls on implementation

1. Spend a few minutes to understand the basic research question / overall objective
2. Pair the research question with the ideal design
3. Ask yourself, do we even need fancy ML ?
4. Make sure your team includes a non-data person who has content expertise
5. Spend time with your end-user to understand their day to day workflow

Implementing machine learning based solutions into real-life: Everything you need to know in 29 minutes

Mike Fralick, MD, PhD, SM, FRCPC

General Internist, Mount Sinai Hospital, University of Toronto

Assistant Professor, Department of Medicine, University of Toronto

@Fralickmike

Machine Learning Course, University of Toronto

20 Oct 2021

Note:
1. α is a smoothing parameter for the level equation ($0 \leq \alpha \leq 1$). The one-step-ahead forecast for time T + 1 is a weighted average of all of the observations in the series $y_1, \ldots, y_T$. If α is small, more weight is given to observations from the more distant past. If α is large, more weight is given to the more recent observations.
2. γ (similar to α) is a smoothing parameter for the seasonal equation ($0 < \gamma < 1 - \alpha$) 3. “ANA” stands for additive error, no trend, additive seasonality