Machine Learning for Healthcare

Deep Learning Indaba 2018, Stellenbosch

Konstantina Palla

@konstantina_pll

Agenda - What we will be discussing

Machine Learning

Data

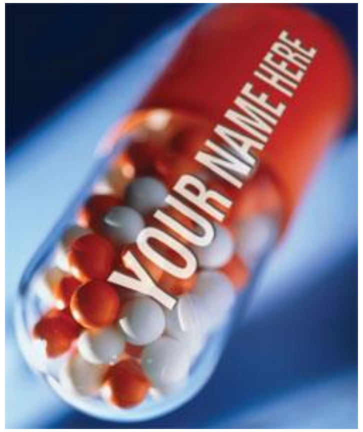
topics

- Uncertainty - Probabilistic Reasoning
- model-free and model-based approaches
- Causality
- mHealth framework
- Other Challenges

Personalised Healthcare

Prognosis, Diagnosis, Treatment… tailored to the individual

What is Personalised Healthcare



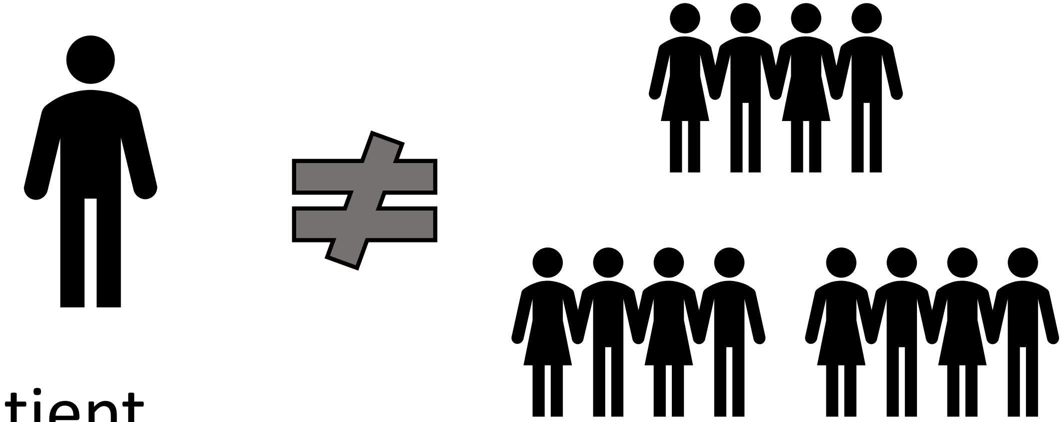
- Traditionally -> *personalised medicine*

"use of individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease."

[National Human Genome Research Institute]

But genomics is not enough

" Need to understand the patient condition, its dynamics and provide optimal patient treatment."



Patient

Factors of disease heterogeneity:

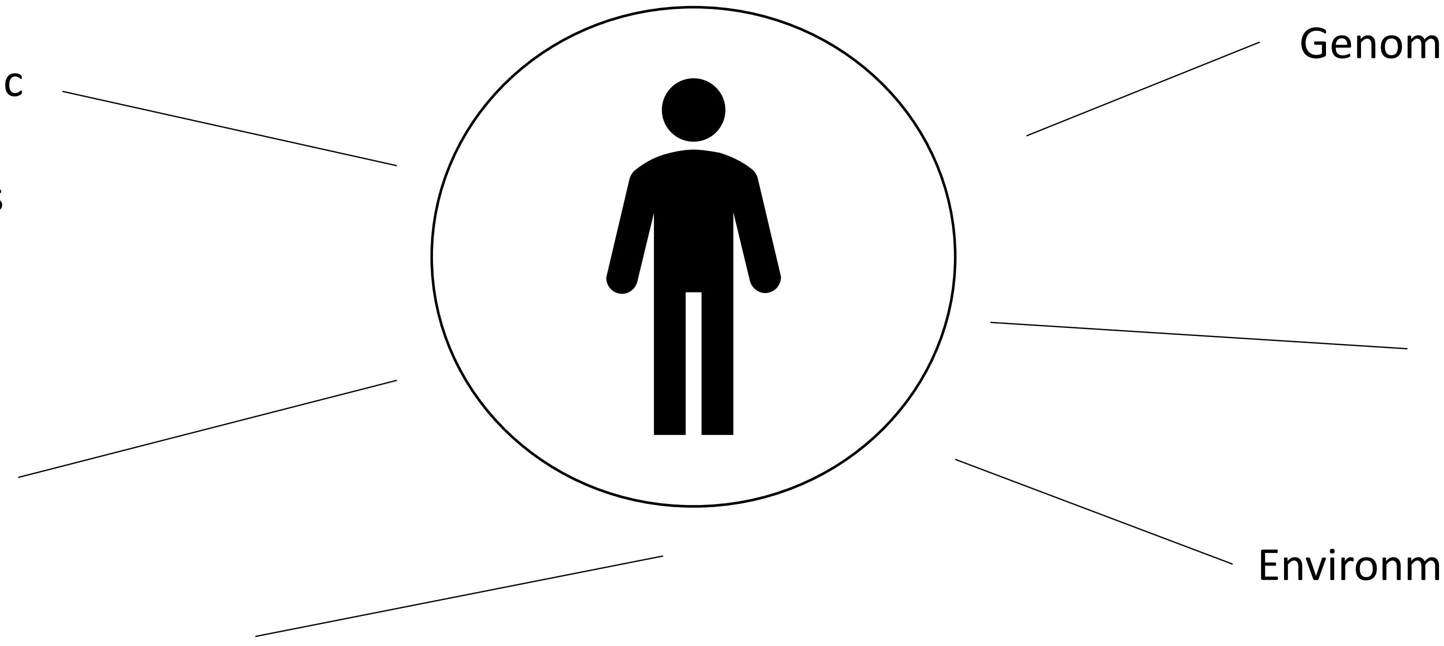
Genomics
Behaviour
Prior exposures
Comorbidities
etc.

We need to be able to capture this variability
→ individualised support provision

What is Personalised Healthcare

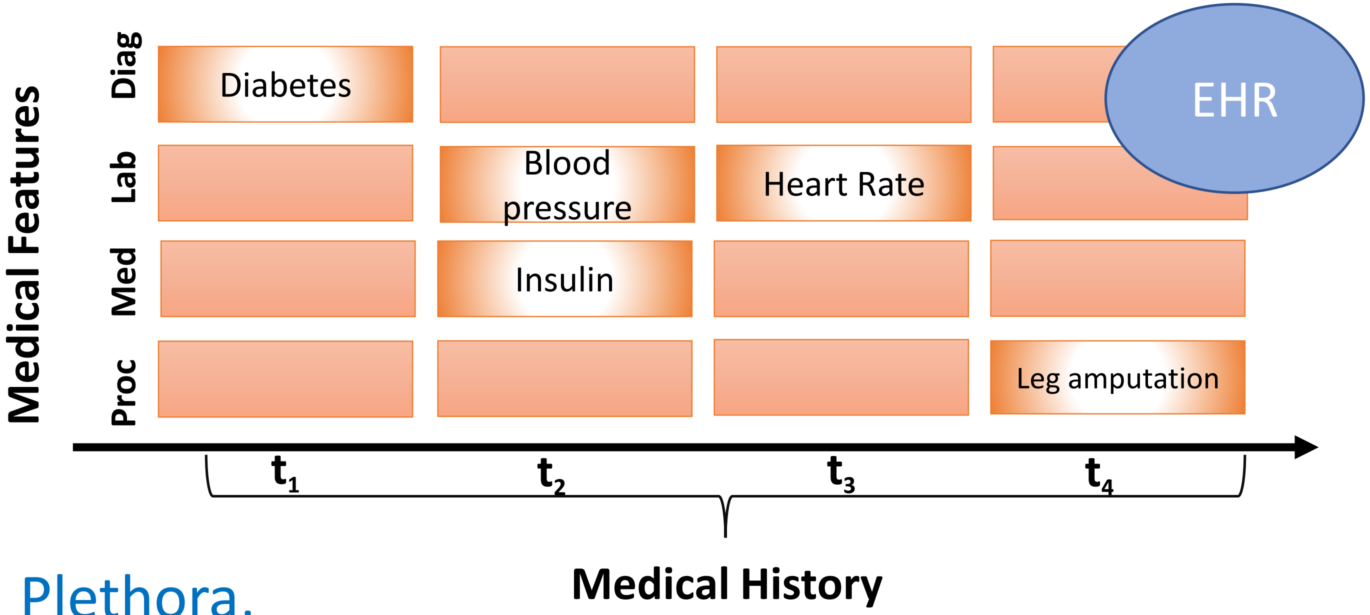
Person in the centre.

Person as unique
individual.



Provision of Prognosis, Diagnosis, Treatment **tailored** to the individual

Health Data – information



Accelerometer
GPS
Gyroscope
Magnetometer
Microphone

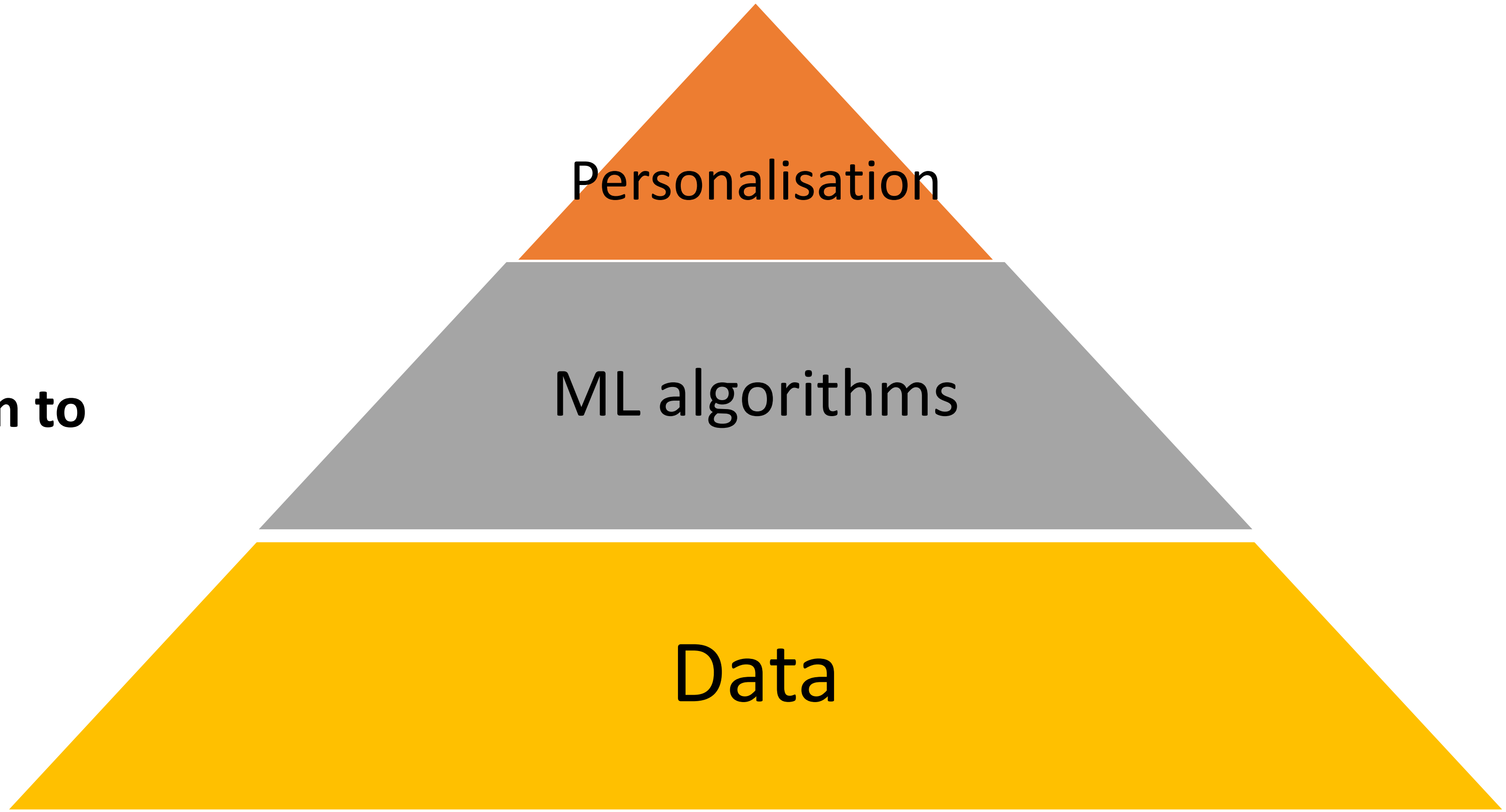
Plethora, multimodal information

How can we extract useful knowledge?

ML can transform data into actionable information

Inspired by **Lee Chonhoo, Luo Zhaojing, Ngiam Kee Yuan, Zhang Meihui, Zheng Kaiping, Chen Gang, Ooi Beng Chin and Yip Wei Luen James**, 2017.*Big Healthcare Data analytics: Challenges and Applications.* Appears in Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications, 11-41.

Healthcare personalisation as a three level process



What algorithm to choose?

ML for Healthcare

❖ Account for uncertainty

ML for Healthcare

❖ <u>Account for uncertainty</u>

- Disease mechanism unknown
  - Symptoms<>disease ?
- Lack of complete knowledge of patient status
  - Unknown comorbidities
- Uncertainty in data
  - Noisy measurements, etc.

Inherent in the clinical domain

Uncertainty should be taken into account

ML for Healthcare

❖ <u>Account for uncertainty</u>

Inherent in the clinical domain ⚠

➢Patient has symptoms of rare disease – 0.2% prevalence

$\mathrm{P}(D) = 0.002$ **prior belief (pretest)**

➢A screening test with blood sample costing 250$ is available

$\mathrm{P}(T|D) = 0.85 \quad P(T) = 0.08$

Decision making under uncertainty
→ need for a principled framework

➢ ? Shall the doctor recommend screening test?

$\mathrm{P}(D|T) = ?$ **posterior belief (posttest)**

Probabilistic Reasoning

Use of **probability** to quantify uncertainty

Bayes' Rule

Probabilistic Reasoning

Bayes' Rule: Update the belief under the light of information

Posterior: Our updated belief after getting more information

likelihood

Bayes' Rule

$$\mathrm{P}(D|T) = \frac{P(T|D)P(D)}{P(T)}$$

prior

posterior

Remember: $\mathrm{P}(D) = 0.002$

$$\mathrm{P}(D|T) = ?$$

$$\mathrm{P}(D|T) = \frac{0.85 * 0.002}{0.08} = 0.021$$

ML for Healthcare

1. Account for uncertainty
2. Fit **and** improve our understanding

“ Need to *understand the patient condition*, its dynamics.”

ML for Healthcare

Fit **and** improve our understanding

" Need to *understand the patient condition*, its dynamics."

<u>Example</u>: Asthma.

Variability in several domains; symptoms,
disease pattern over time, response to treatment etc.

Exact cause is **unknown**.

#facts

ML for Healthcare

Fit **and** improve our understanding

" Need to *understand the patient condition*, its dynamics."

Example: Asthma.

Most likely associated with a combination of genetic (inherited) and environmental factors (such as allergens and infections). **#assumption**

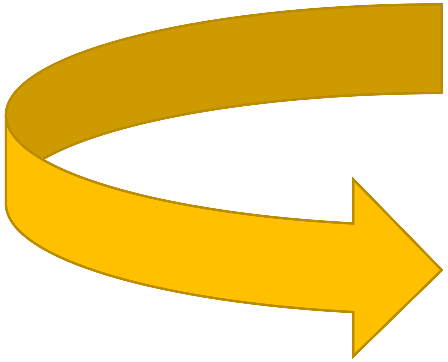
Approach A:
Assume nothing, let the data speak

model-free versus model-based approaches

Approach B:
We have a hypothesis/assumption of the mechanism.

ML for Healthcare

“ Need to *understand the patient condition*, its dynamics”



Model – free approaches

Model-free approaches

Adapt to the intrinsic data characteristics

No (or few) assumptions -- > they don't explain how the data was generated.

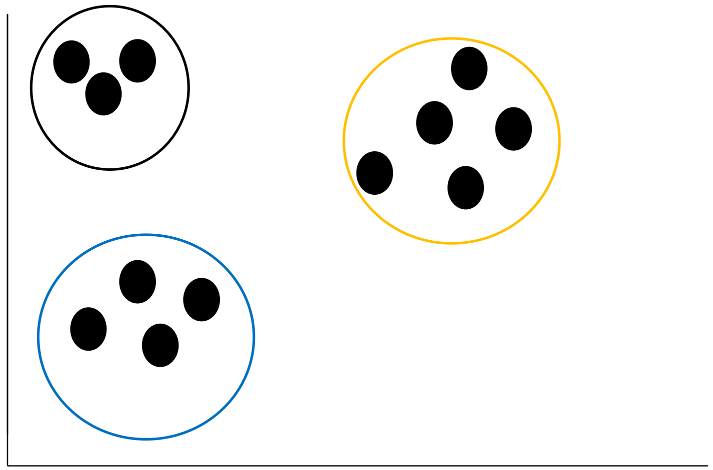
Decision Trees	K-means	Neural Networks	
Random Forests	Nearest Neighbour	Support Vector Machines	
Ensemble Methods	Hierarchical Clustering	Regression	. . .

<u>Choice:</u>

- ❑ As a first step towards understanding
- ❑ Familiarity of the user with the algorithm
- ❑ Availability of the corresponding software implementation

Model-free approaches - Example

Clustering



Dot→ patient info
[coughing, wheeze, age of onset, asthma duration, lung function,..]

+ They force the pattern to be captured
- They don't explain was the data was generated
  - Focus on the data, not on the process

Model-free approaches – Application on ASD

Autism Spectrum disorders (ASDs): a developmental disorder that affects communication and behaviour.

Social communication difficulties

Restricted interests

Repetitive behaviours

Spectrum: Wide variation in the type and severity of symptoms (**heterogeneity**)

- Motivation: Classifying patients into similar groups would provide a powerful tool to individualise treatment regimes

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Model-free approaches – Application on ASD

- ASD and Comorbidities

A disease or a syndrome that co-occurs with the target disease

Gastrotestinal disorders

Epilepsy

Sleep disorders

Muscular dystrophy

Psychiatric illnesses

…

Investigate the patterns of co-occurrence of medical comorbidities in ASDs.

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Model-free approaches – Application on ASD

Patients: ~ 5K Children

Data: Comorbidity counts over period 0-15year split in 6month window and for 45 comorbidities.

Method: Unsupervised clustering

D = 1350

Patient vector [_,_,…,_,_|_,_,…,_,_|. . . _,_,…,_,_| _,_,…,_,_]

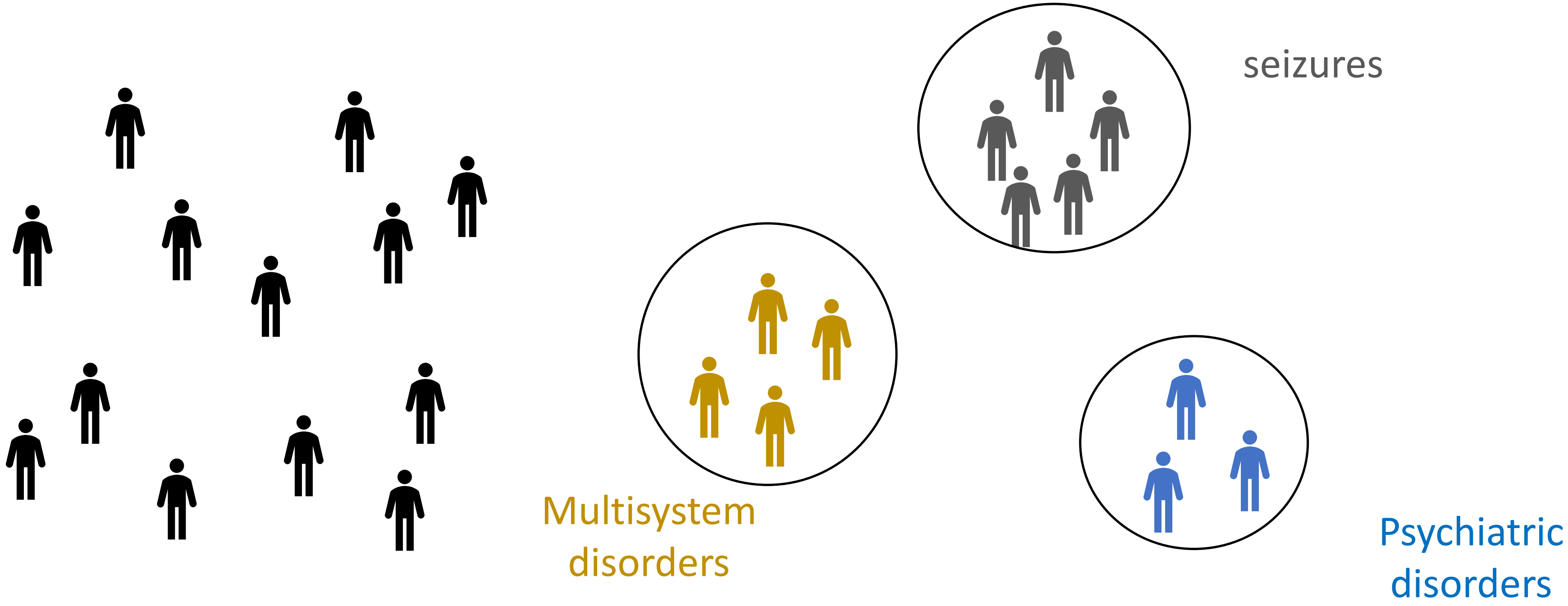
45 comorbidities

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Model-free approach – Application on ASD

Results:

Three distinct subgroups were identified



Better understanding of co-occurrence of comorbidities in ASDs

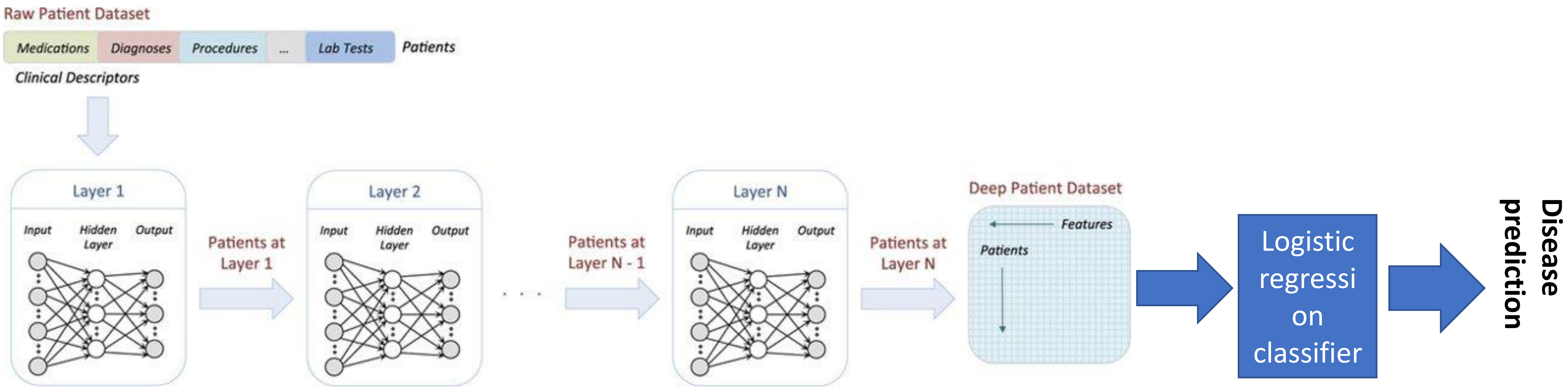
A first step for uncovering underlying etiologies

Similar work on Diabetes type 2 by [**Ahlqvist et al, 2018]**

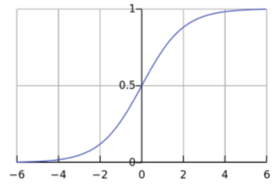
Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Model-free – Example Application on Disease prediction

Disease Prediction using Deep Neural Networks



Deep feature learning method
Derives a general-purpose patient representation



R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "***Deep Patient****: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records.*" Scientific reports, vol. 6, no. April, p. 26094, 2016.

ML for Healthcare personalisation

<u>Example</u>: Asthma.

Most likely associated with a combination of genetic (inherited) and environmental factors (such as allergens and infections). **#assumption**

Approach B:
We have a hypothesis/assumption of the mechanism.

Model – based approaches

+ probabilistic framework

Model-based approach

What is a model?

Definition [**Bishop et al.**, 2015]

"A set of assumptions about a problem domain expressed in a precise mathematical form, that is used to create a ML solution"

A set of assumptions (defined by the user) to describe how the observed data is generated.

Model-based approach

A set of assumptions (defined by the user) to describe how the observed data is generated.

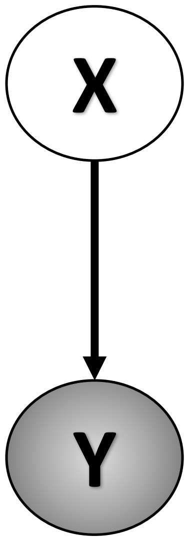
Assumptions

our believes of how the data
is generated
(latent mechanism
responsible for the observations)

(e.g. allergy, genetics, medication etc.)

Observed data
(e.g. clinical findings or in asthma:
Wheeze, rhinitis, lung function etc.)

Graphical model



Model:
set of vars
dependencies

Tailored to the data

One of possibly many

Model based approach - Uncertainty

Uncertainty in many forms

Model

Value of latent parameters

Observations (noise)

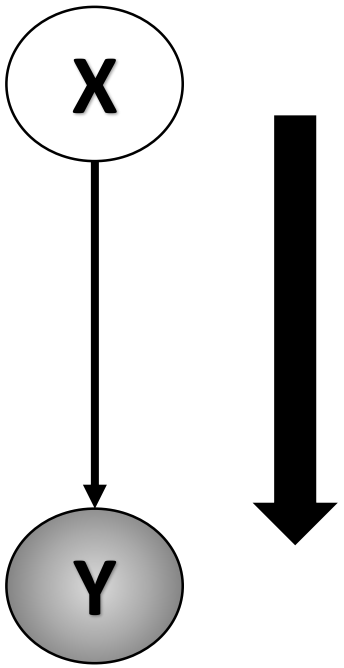
Probability theory to express all forms of uncertainty

Model based approach - Uncertainty

Probabilistic model

Probability distributions to represent all the uncertain unobserved quantities

Prior belief p(x)



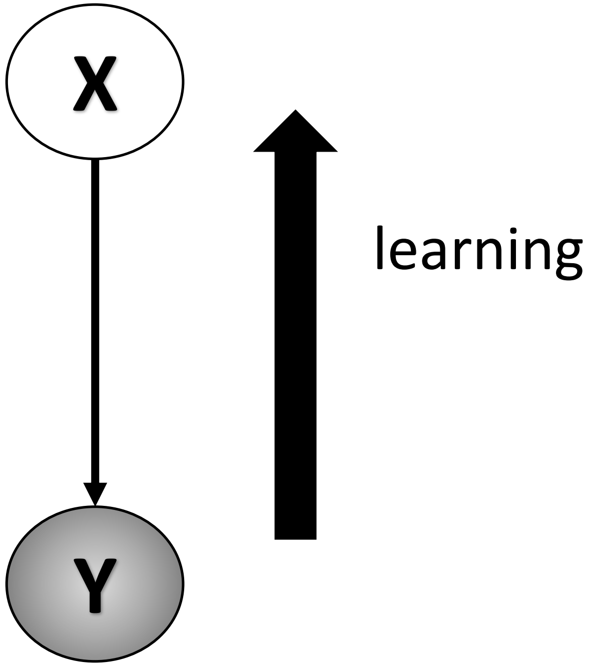
Generative process

$$x \sim p(x)$$
$$y \sim p(y|x)$$

and how they relate to the data

p(y|x)

Model based approach - Learning



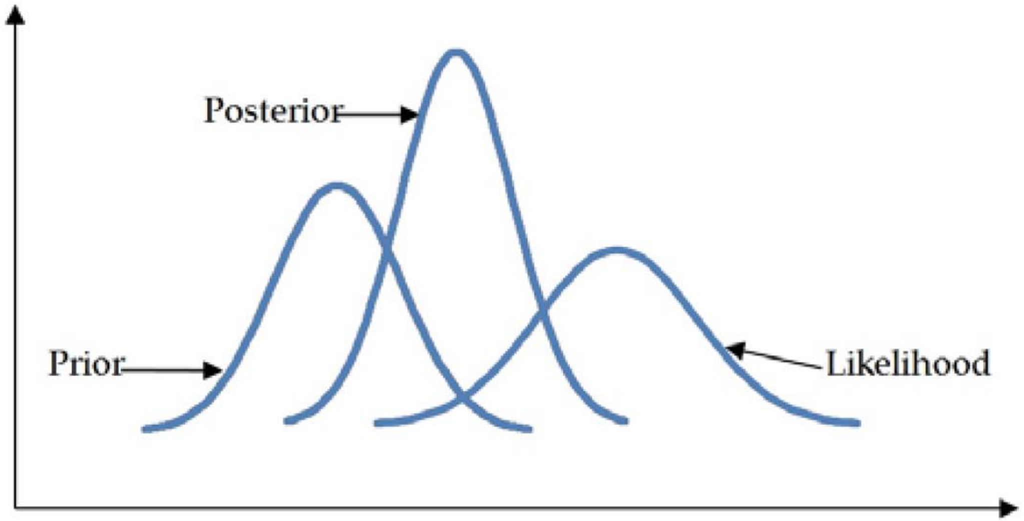
Learning: infer the value of the unknown quantities.

Posterior: Our updated belief after having seen the data

Bayes' Rule

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(x|y) \propto p(y|x)p(x)$$



Model based approach - Example

- Motivation:

Heterogeneity in complex diseases (chronic).
Scleroderma.

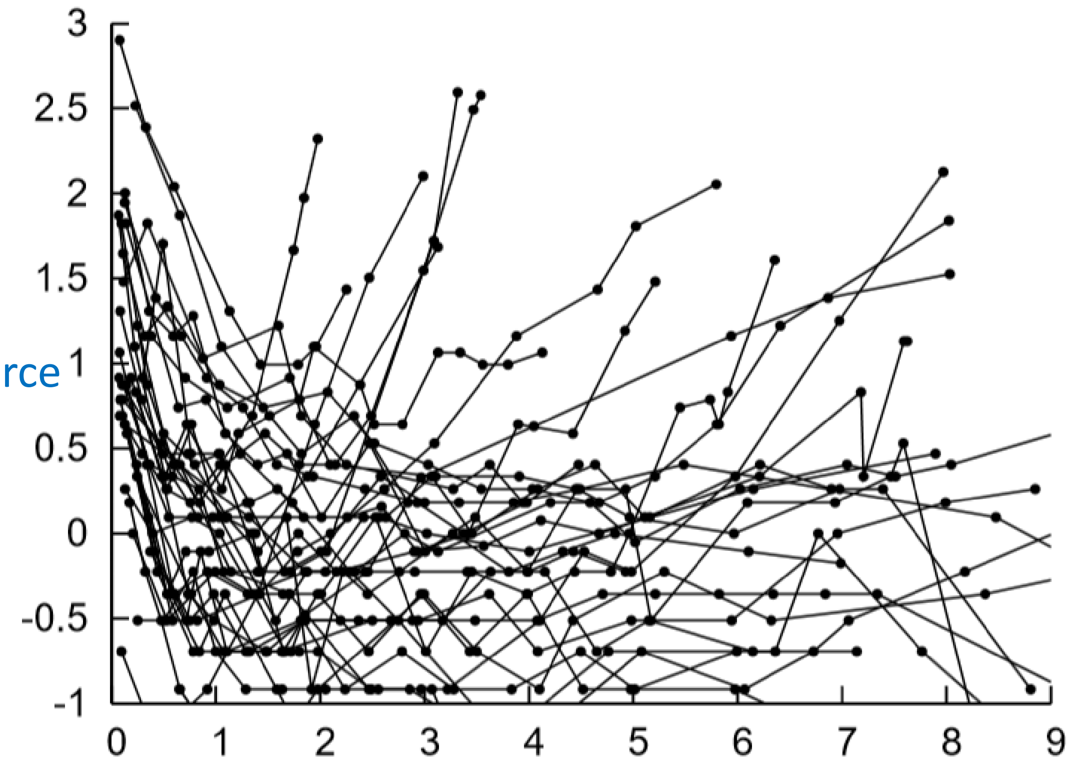
- Target:

Predict future disease trajectory

Trajectory of lung severity over time (PFVC-predicted force vital capacity)

- Challenge:

Individualize prediction by capturing variability

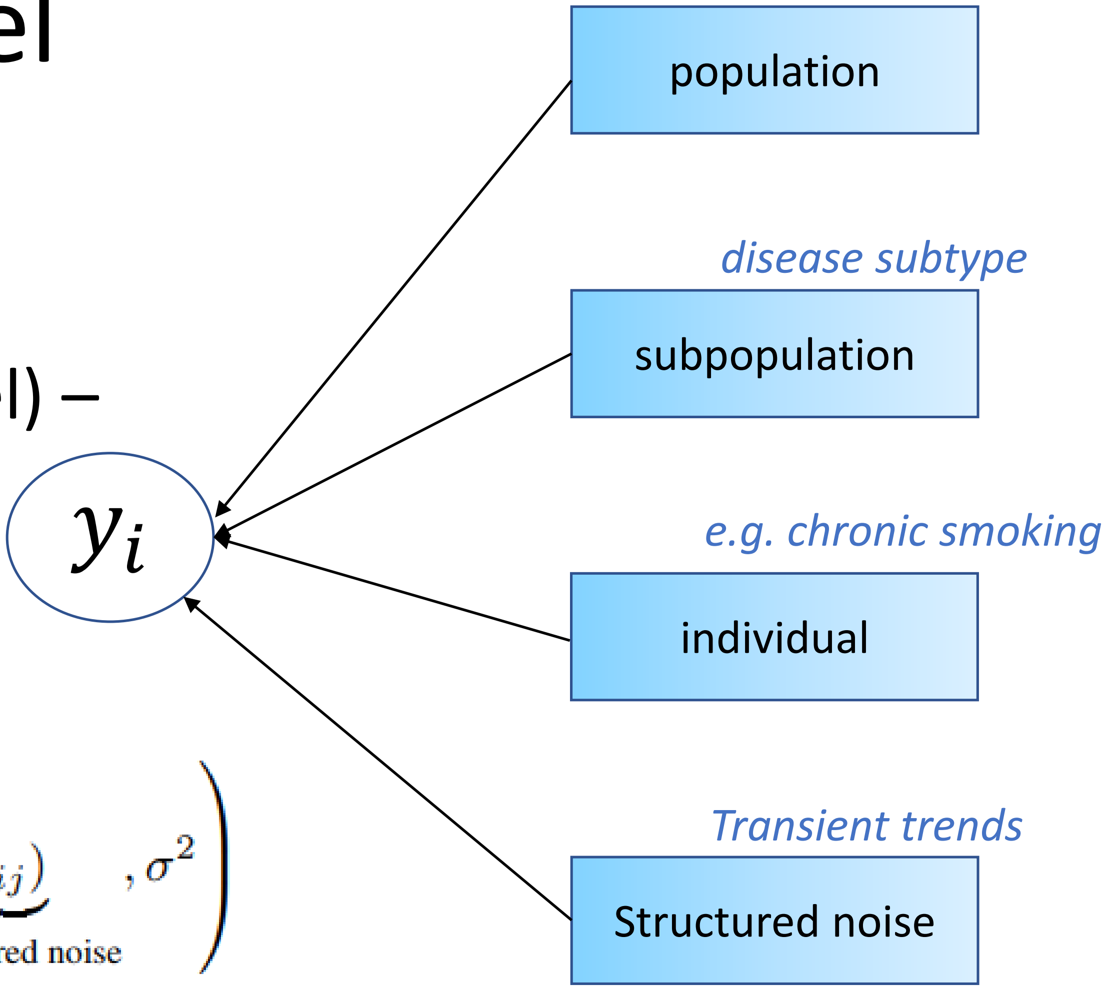


P. Schulam, S. Saria. ***A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-resolution Structure.*** Neural Information Processing Systems (NIPS), 2015.

Model based approach – Individualised disease progression model

- Assumptions: 4 factors of variability
- Model:

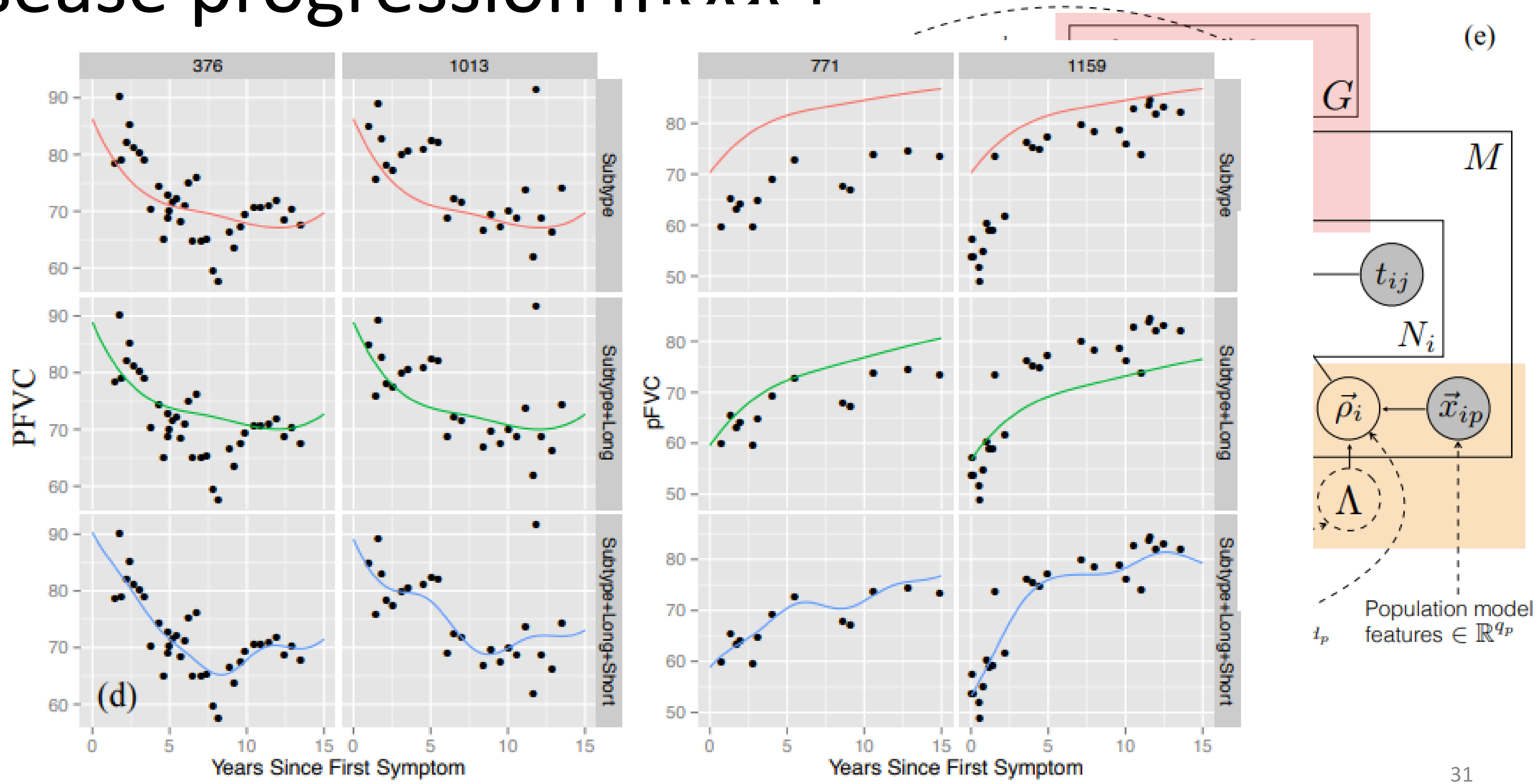
Multi-level model (Latent variable model) – organise variability in different levels



$$y_{ij} \sim \mathcal{N}\left(\underbrace{\Phi_p(t_{ij})^\top \Lambda\, \vec{x}_{ip}}_{\text{(A) population}} + \underbrace{\Phi_z(t_{ij})^\top \vec{\beta}_{z_i}}_{\text{(B) subpopulation}} + \underbrace{\Phi_\ell(t_{ij})^\top \vec{b}_i}_{\text{(C) individual}} + \underbrace{f_i(t_{ij})}_{\text{(D) structured noise}}, \sigma^2\right)$$

P. Schulam, S. Saria. ***A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-resolution Structure.*** Neural Information Processing Systems (NIPS), 2015.

Model based approach – Individualised disease progression model



Model-free Vs Model-based approach

Model-free	Model-based
➢ Learn pattern in the data – no assumptions ➢ Give insight – can be used as first step ➢ Easy to use – off the shelve ➢ Hard to match the requirements of a new application.	➢ Model assumptions ➢ Allow for human-led exploration. ➢ Perfect fit for probabilistic framework – uncertainty ➢ Try many different models to find the best

- Questions?

Causal reasoning in Healthcare

What treatment should I give to patient?

Ideally, we want to be confident answering this.

Drug
- **A**
- **B**

Rephrase:
We are interested in the effect of treatment A and B on the patient.

Causal Reasoning

Causal reasoning in Healthcare

Generative model

$p(X)$ X

association

$$p(Y|X) = \frac{p(X,Y)}{p(X)}$$ Y

observational

Indicates statistical dependence among the variables

Causal model

intervention

X=x

X

causality

$p(Y|do(x))$ Y

interventional

Indicates causal effect

Aren't they the same?

Causal reasoning in Healthcare

Y: the pressure in the espresso machine
X: the reading of the built-in barometer

Generative model

$p(X)$ X 

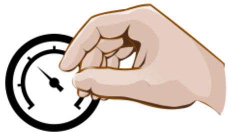
association

$p(Y|X)$ Y

intervention

Causal model

X=x

X 

causality

$p(Y|do(x))$ Y

People who take this treatment are likely to get better

This treatment makes people get better

Which one do I want?

Example borrowed from Ferenc's Huszar blog https://www.inference.vc
@fhuszar

Causal reasoning in Healthcare

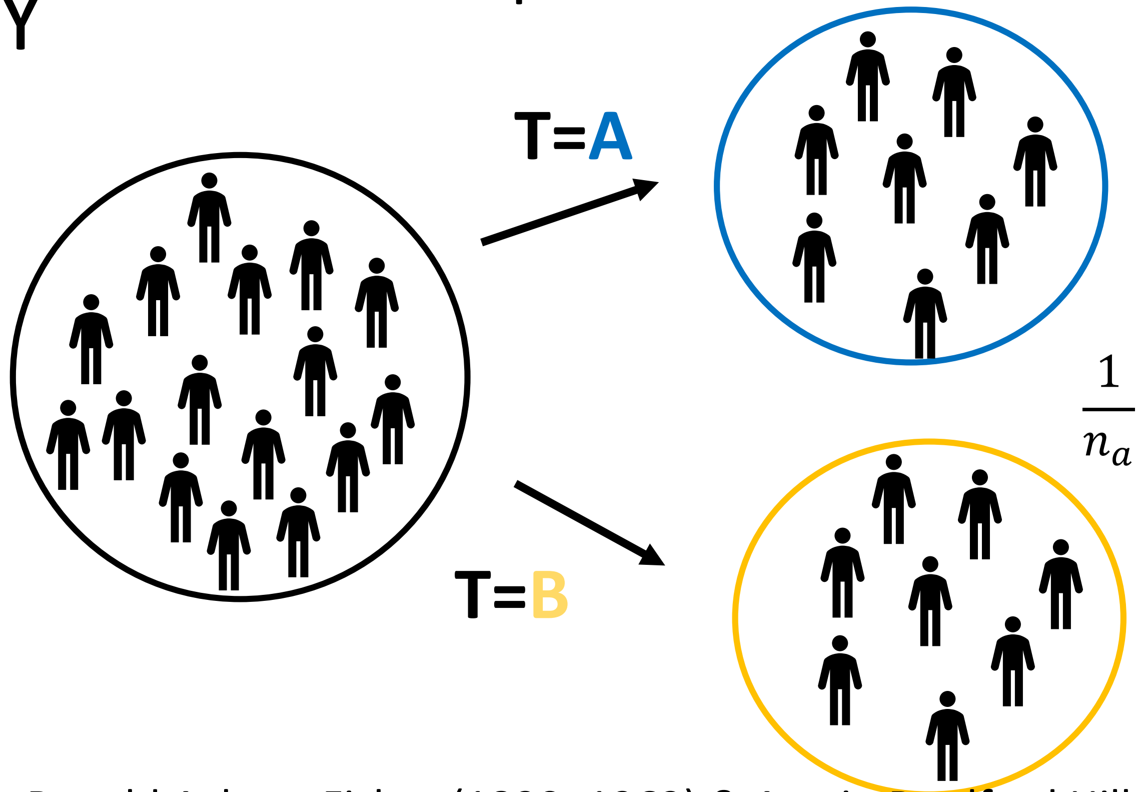
T

$\downarrow$

Y

<u>Randomized</u> Control Trials **"Gold standard"**

Control & Manipulation



Evaluate average treatment effect

$$\frac{1}{n_a}\sum_{i=1}^{n_a} Y_{ia} - \frac{1}{n_b}\sum_{j=1}^{n_b} Y_{jb} = \tilde{Y}_a - \tilde{Y}_b$$

BUT:

- Impractical
  - Expensive (recruiting is hard!)
  - Take time
- Unethical
  - Does inhaling asbestos cause cancer?

Ronald Aylmer Fisher (1890–1962) & Austin Bradford Hill (1897–1991)

Causal reasoning with Observational Data

[Absent controlled experiment, observational data are used]

Cheaper, Faster, in Plethora

Example:

Observations: {Yi, Ti}

Response of individuals to a treatment

1

2

Treatment

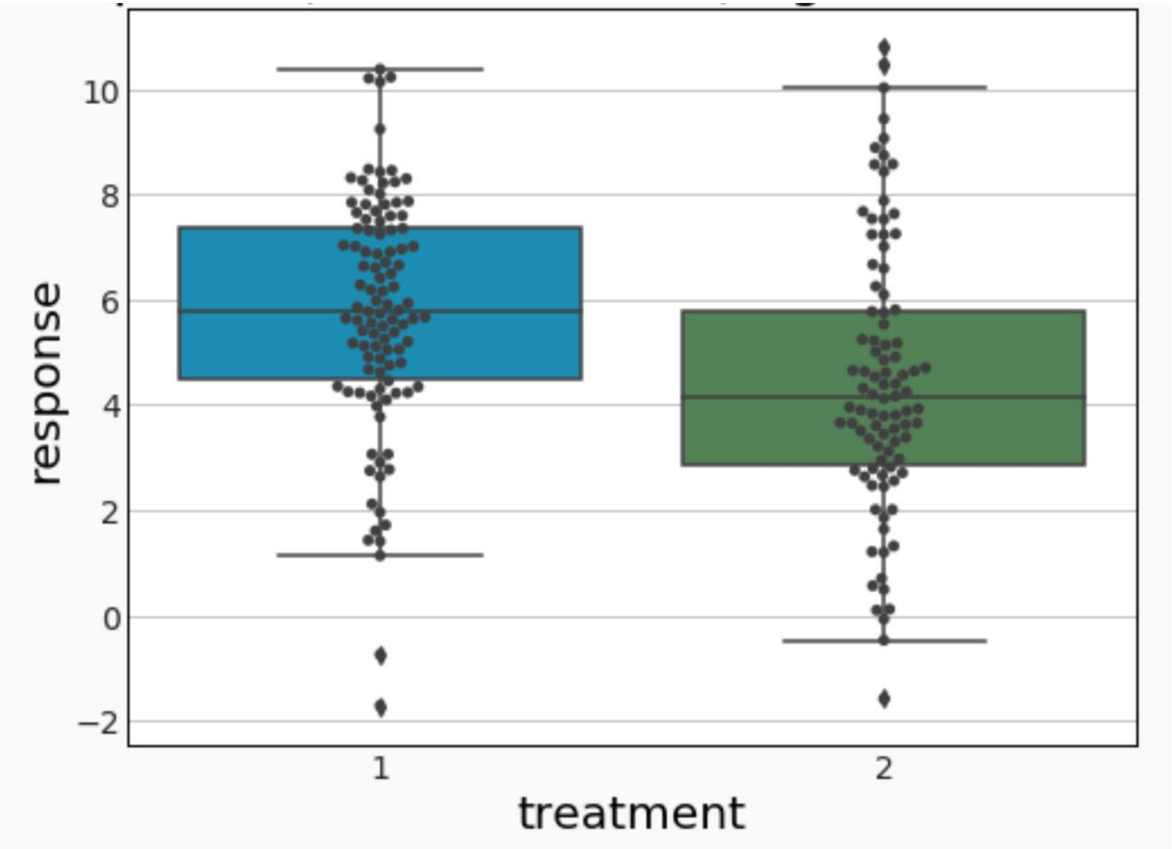
↓

Response

Can we just “emulate” a RCT?

Causal Reasoning with Observational Data

A lower response is better.



"Treatment 2 is a more effective treatment" ?

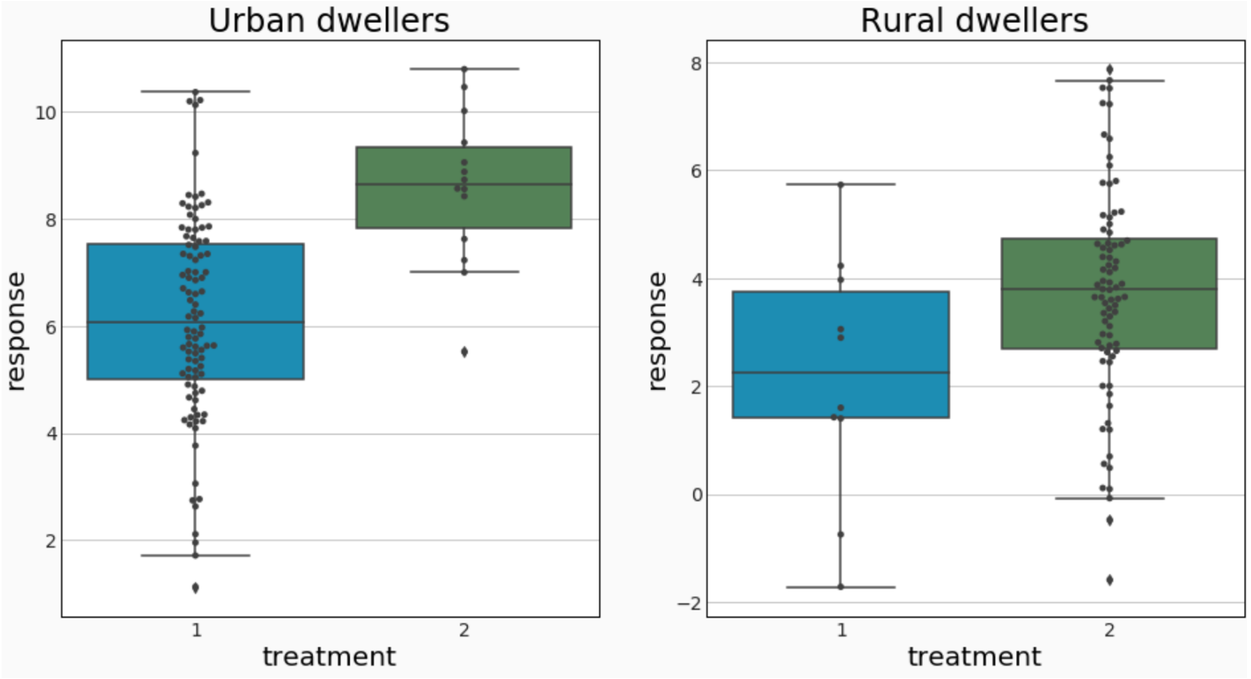
Treatment → **Response**

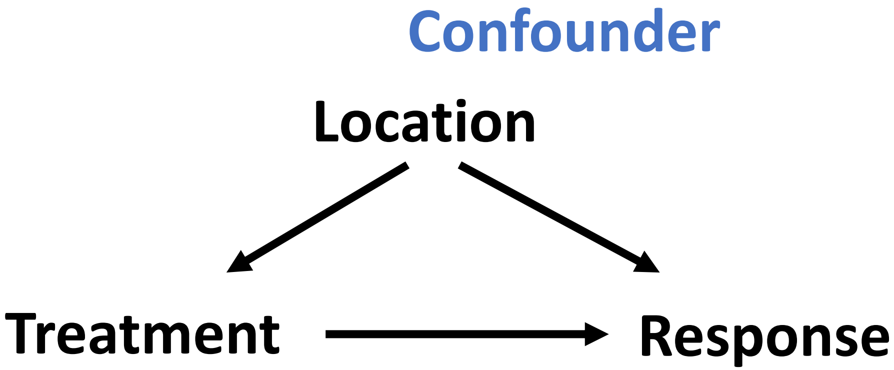
Causal reasoning with Observational Data

Limitations

- Unknown underlying data collection protocol - Confounders
- Doesn't contain all possible outcomes for all treatments for a pat[illegible]nt

Causal Reasoning with Observational Data - confounding





confounder is a variable that causally effects both the (intervention var) treatment and the outcome.

Causal Reasoning with Observational Data - confounding

Solution: control for the effects of the confounders

- Identify them and study groups of similar individuals separately
- Propensity scores method
  - modify your analysis so that the covariates (say age, sex, gender, health status) were "balanced" between the treatment groups
- Regression-based methods

- Other, e.g. *do*-calculus method, structural equation models (SEMs) etc.

Causal Reasoning with Observational Data - confounding

Solution: control for the effects of the confounders

➢Regression-based methods, example

$$\mathrm{Y_i} = \beta_0 + \beta_1 T_i + \beta_0 x_i + \epsilon_i$$

1 0

$\beta_1 = (Y_i|T_i = 1) - (Y_i|T_i = 0)$ treatment effect

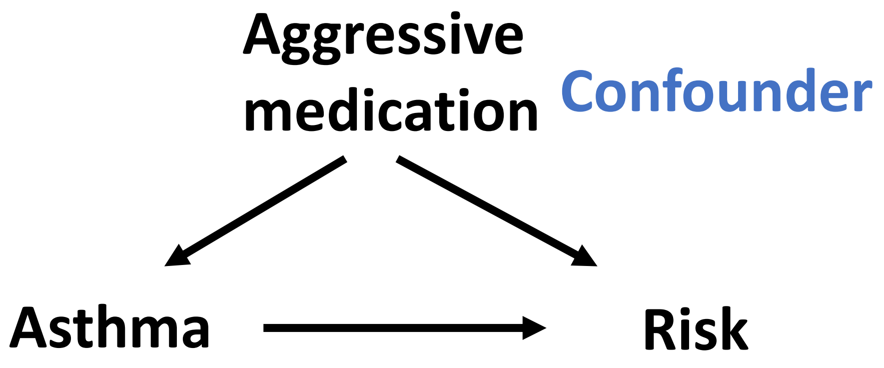
- We can estimate a causal effect in regression if:
  - the regression model includes all confounders; and
  - the regression model is correct (linearity in the response space)

Causal Reasoning with Observational Data

Pneumonia example [**Caruana et al.**, 2015]

Machine Learning to guide the treatment of pneumonia patients

What the model inferred: Asthmatic patients have less risk of death!



Causal reasoning with Observational Data

Limitations

- Unknown underlying data collection protocol - Confounders
- Doesn’t contain all possible outcomes for all treatments for a patient

ML for personalised Treatment - Counterfactuals

Problem: Evaluate individual Treatment effects using observational data

Assume: $Y_i^{(A)}, Y_i^{(B)}$ outcome after the patient i is given treatment {A, B}.

Challenge:

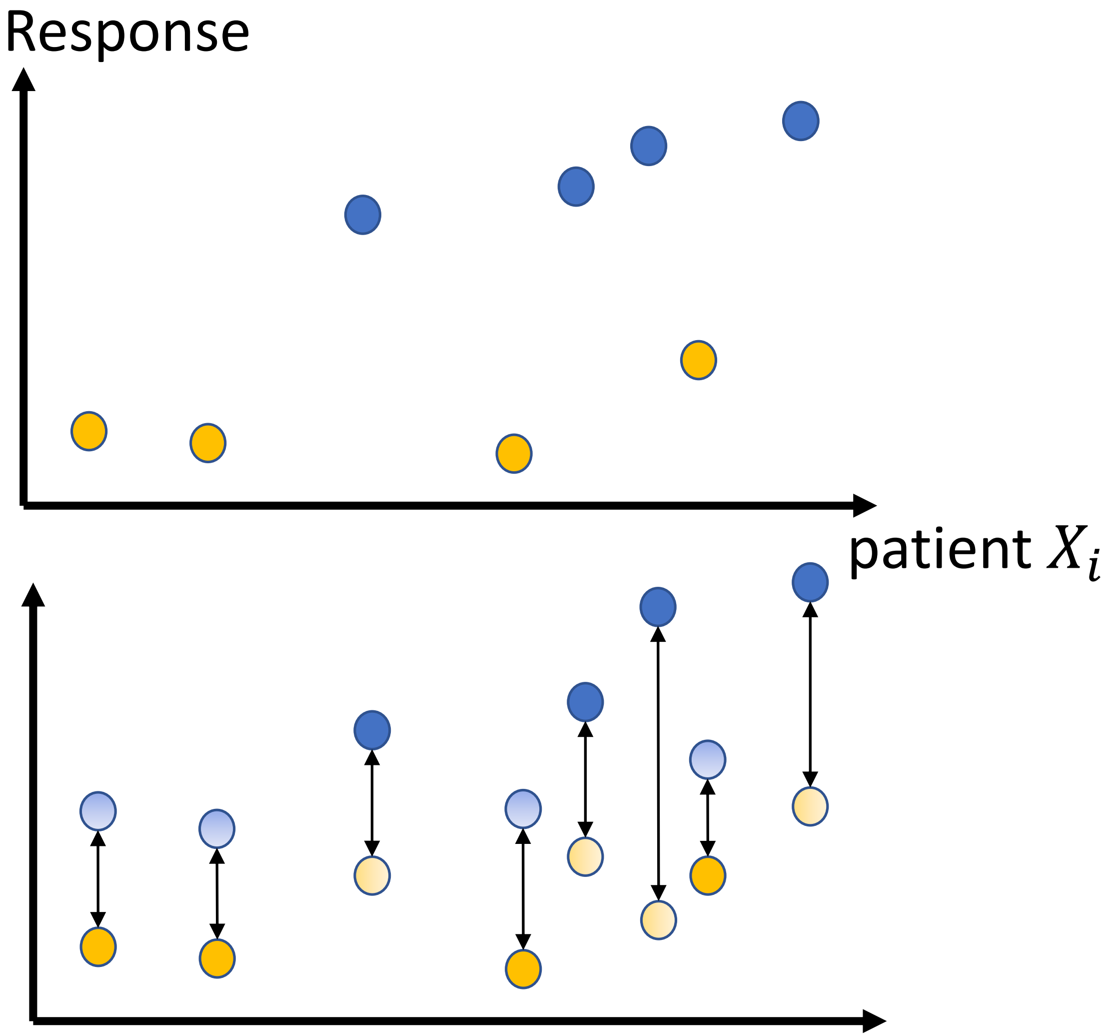
- Evaluate Treatment effect for a patient $Y_i^{(A)} - Y_i^{(B)}$ using **observational data** – “What if?”
- BUT: For every subject we only observe one outcome FACTUAL

Never observe the counterfactual.

Observed patient response to treatment A

What would the outcome be if the patient was given treatment B?

ML for personalised Treatment - Counterfactuals



Idea: Compute distribution over counterfactuals.

Ahmed M. Alaa and Mihaela van der Schaar. *Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes*, NIPS, 2017.

Frederik D. Johansson, Uri Shalit and David Sontag. *Learning Representations for Counterfactual Inference. ICML, 2016.*

ML for Healthcare

ML for mHealth

Mobile Health and personalised interventions

Accelerometer

GPS

Gyroscope

Magnetometer

Microphone

Machine Learning

Actionable information (intervene)

Improve health

personalised

Mobile Health and personalised interventions

- **Intervention app - Fundamental pattern that repeats over time**

1. at a given time point do
2. mobile phone collects data (the context)
3. a decision rule (or policy) maps the data into an intervention option (the action)
4. mobile phone records the outcome (interpreted as a reward, so higher is better)
5. done

GPS
accelerometer
Agenda
Weather etc.

Intervention options:
Text messages for walking
Going to the gym
Summary of past workouts etc.

Minutes of activity

Mobile Health and personalised interventions

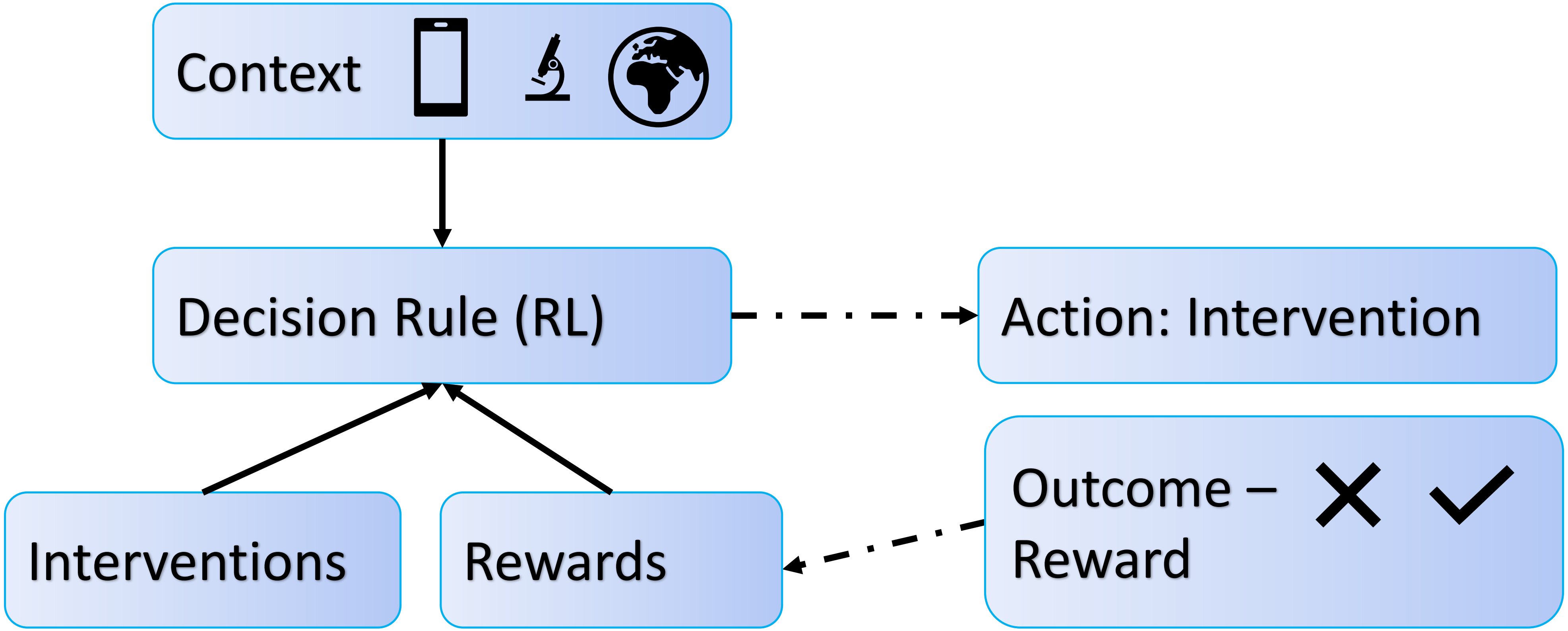
a decision rule (or policy) maps the context into an intervention option (the action)

Reinforcement learning framework + contextual bandits

Exploration - Exploitation

Personalised action

Mobile Health and personalised interventions



Mobile Health and personalised interventions

Example:

An intervention app to encourage physical activity in Diabetes patients

Approach: Encourage physical activity through personalised messages

Method: RL with contextual bandits

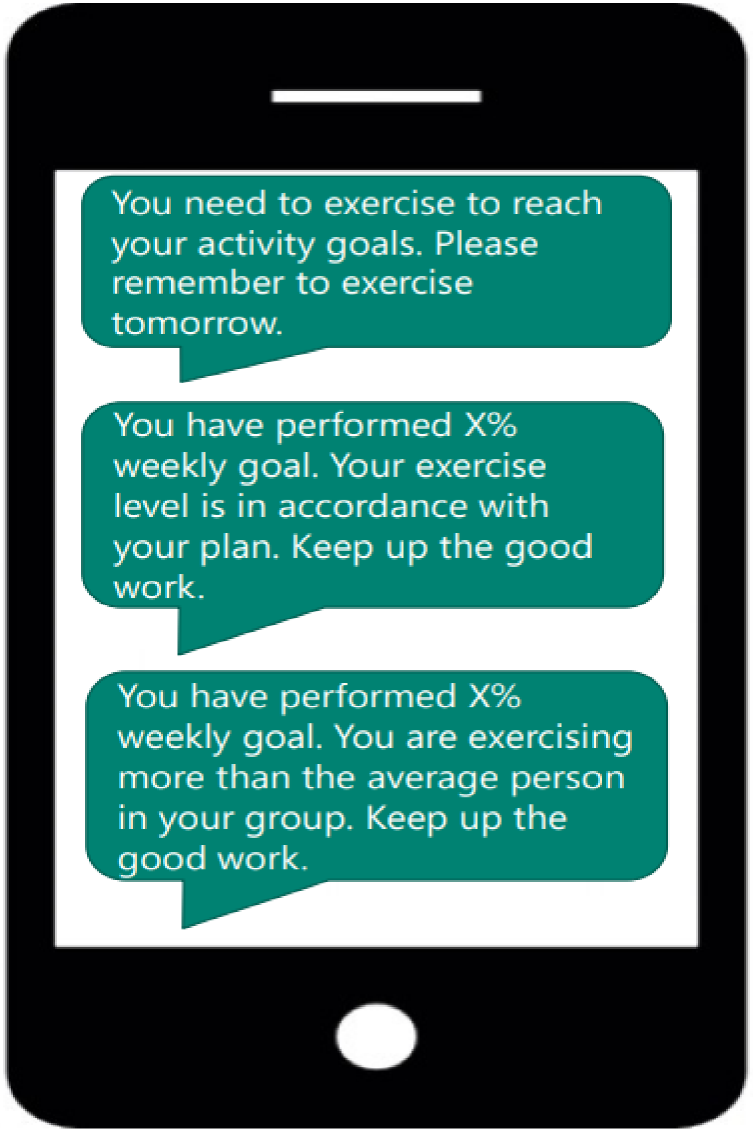
Irit Hochberg, Guy Feraru, Mark Kozdoba, Shie Manor, Moshe Tennenholtz, EladYom-Tov. *Encouraging Physical Activity in Diabetes Patients Through Automatic Personalized Feedback Via Reinforcement Learning Improves Glycemic Contro*l. Diabetes Care Jan 2016.

Mobile Health and personalised interventions

Negative feedback

Positive feedback relative to self

Positive feedback relative to others



Slide by **Elad Yom -Tov**

Mobile Health and personalised interventions

minutes of activity in the last day
Cumulative number of minutes of activity this week
Fraction of activity goal

Age
Gender

Allows information from past (historical)

Each person modelled through several aspects

Personalised messages

User vector X

Augment with action vector A

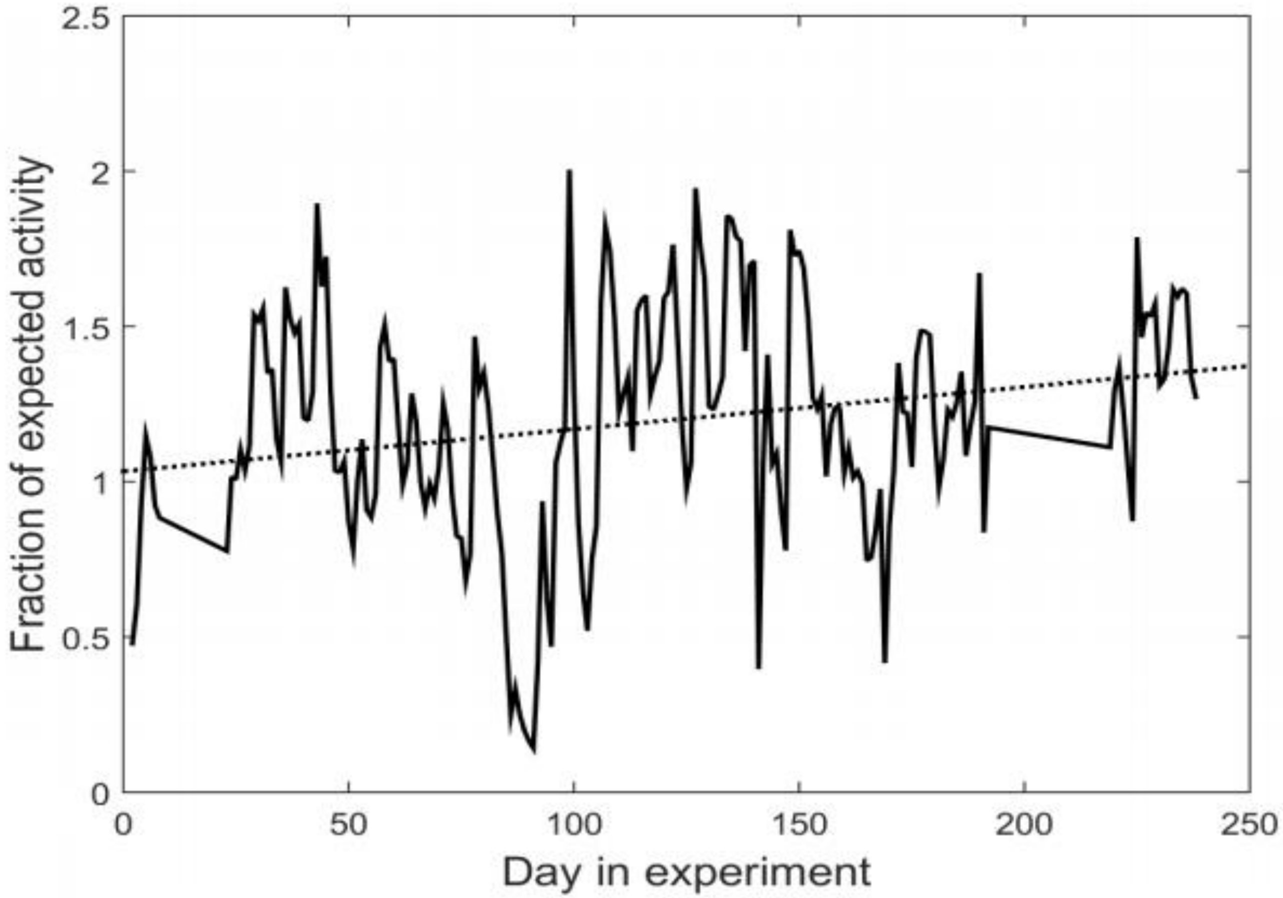
Predict respective rewards Y of [X, A]

Exploit and explore – choose message (action)

Send message

Mobile Health and personalised interventions

27 patients were recruited for a period of 6 months each, 1/3 served as controls



Mobile Health and personalised interventions

Questions to consider:

When to send the interventions?

- Just-In-Time-Adaptive-Interventions (JITAIs) [**Inbal et al., 2016**]

More than ML science

Need to understand the user

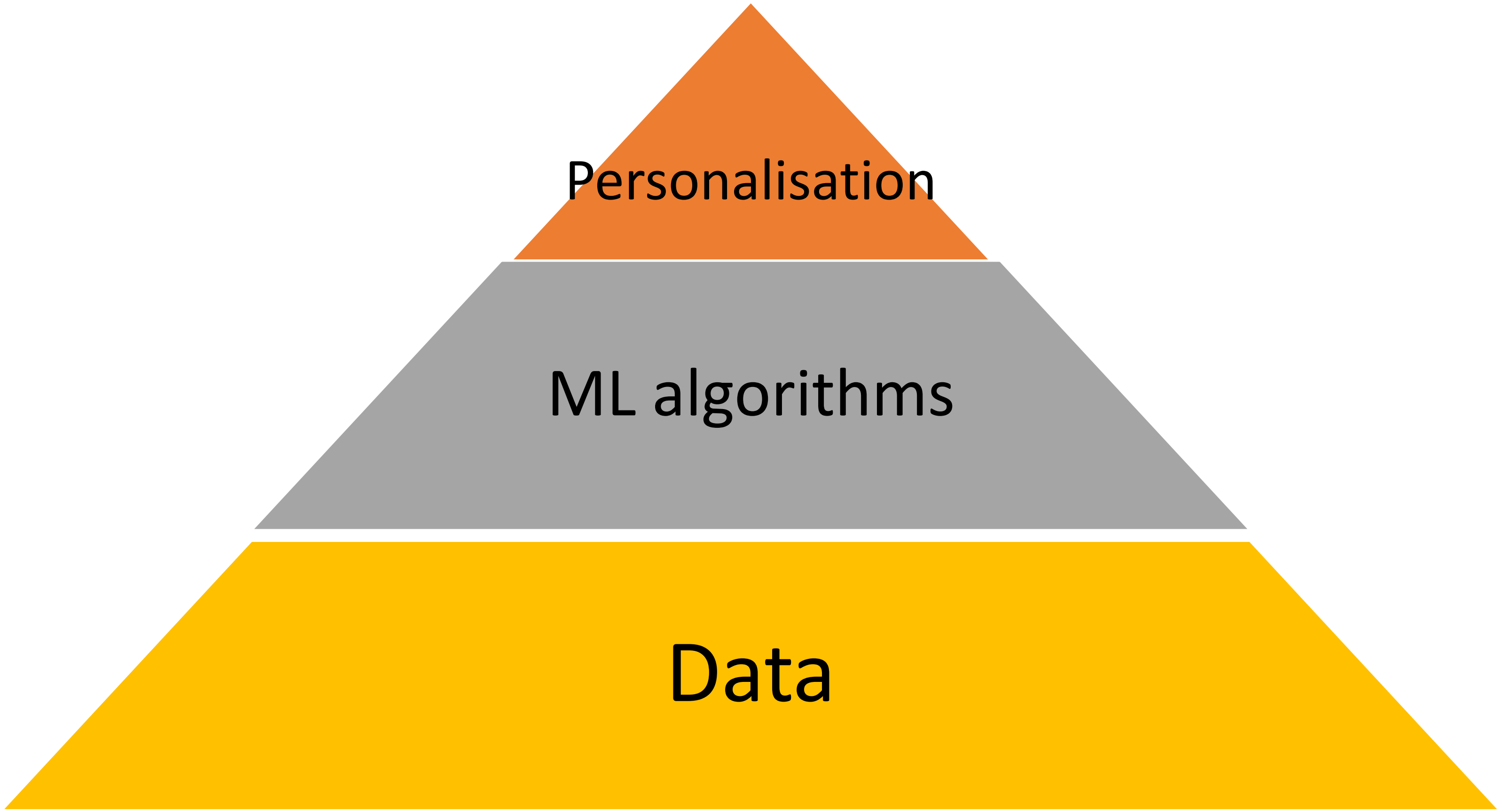
- Psychologists, Behavioural scientists, HCI experts.

Need synergy of sciences

Healthcare personalisation as a three level process

ML algorithms:
Transparency,
interpretability

Data:
Quality
Feature engineering
Data fusion
Missingness



Thank you

Bibliography

Lee Chonhoo, Luo Zhaojing, Ngiam Kee Yuan, Zhang Meihui, Zheng Kaiping, Chen Gang, Ooi Beng Chin and Yip Wei Luen James, 2017.*Big Healthcare Data analytics: Challenges and Applications.* Appears in Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications, 11-41.

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Ahlqvist, Emma et al. *Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables.* The Lancet Diabetes & Endocrinology, 2018, Volume 6 , Issue 5 , 361 - 369

Caruana, Rich & Lou, Yin & Gehrke, Johannes & Koch, Paul & Sturm, Marc & Elhadad, Noemie. (2015). *Intelligible Models for HealthCare.* 1721-1730..

Irit Hochberg, Guy Feraru, Mark Kozdoba, Shie Manor, Moshe Tennenholtz, EladYom-Tov. *Encouraging Physical Activity in Diabetes Patients Through Automatic Personalized Feedback Via Reinforcement Learning Improves Glycemic Contro*l. Diabetes Care Jan 2016.

Bibliography

P. Schulam, S. Saria. ***A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-resolution Structure.*** Neural Information Processing Systems (NIPS), 2015.

Judea Pearl. *Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*. Technical report,University of California, US, 2017.

Ahmed M. Alaa and Mihaela van der Schaar. *Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes*, NIPS, 2017.

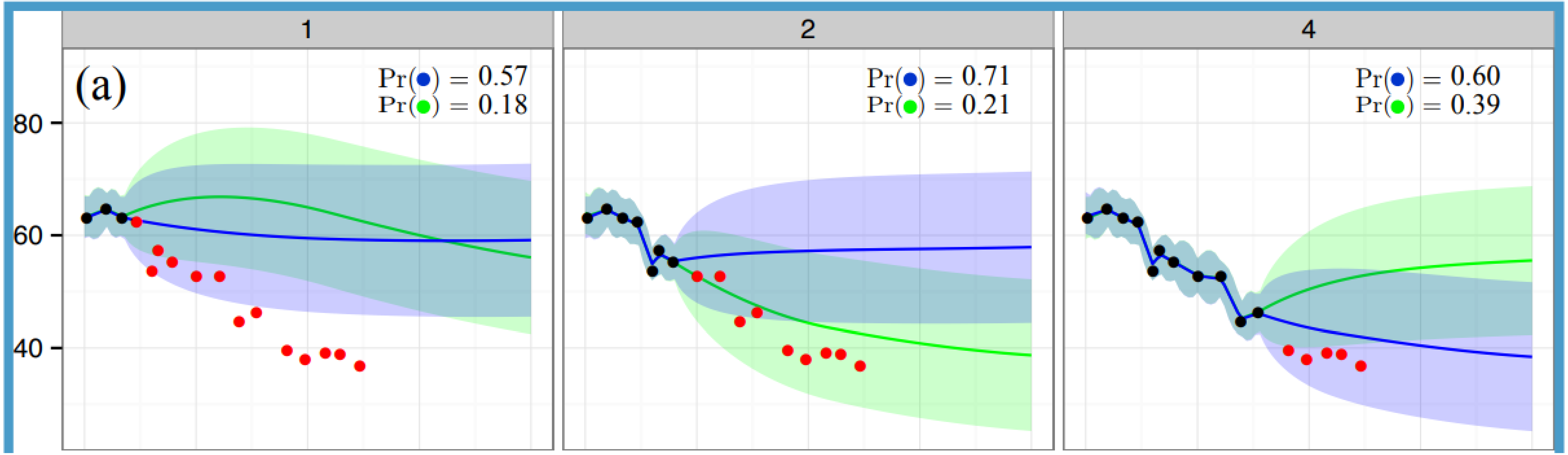
Frederik D. Johansson, Uri Shalit and David Sontag. *Learning Representations for Counterfactual Inference. ICML, 2016.*

Bibliography

Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. *Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support*. Annals of Behavioral Medicine, 2016. accepted subject to revisions.

Bishop, Christopher and Winn, John and Diethe, Tom, 2015. *Model-based Machine Learning.* Retrieved by http://www.mbmlbook.com/

Model based approach – Individualised disease progression model



Model based approach - Learning

Model + Inference = Machine Learning algorithm

Computational process of learning

Flexibility

- Different inference algorithms can be run on the same model

Consistency

- You can create one model and query it in different ways

Maintainability

- If you want to refine the assumptions encoded in the model, the clean separation makes it straightforward to update it.