ICML 2018 TUTORIAL: MACHINE LEARNING FOR PERSONALISED HEALTH

Danielle Belgrave (Microsoft Research Cambridge)
Konstantina Palla (Microsoft Research Cambridge)
Lamiae Azizi (University of Sydney)

TUTORIAL: MACHINE LEARNING FOR PERSONALISED HEALTH

Danielle Belgrave (Microsoft Research Cambridge)

Microsoft

Imperial College London

THE UNIVERSITY OF SYDNEY

Roadmap

- **Part 1** — Introduction: Drivers of machine learning in healthcare
- **Part 2** — Critical Evaluation of data-driven healthcare
- **Part 3** — ML strategies for healthcare personalisation
- **Part 4** — Future challenges of ML for Healthcare

Part 1: What are the drivers of machine learning in healthcare?

Wellness and self-care personalisation: patient perspective

Population data-driven healthcare: policy perspective

Precision drug discovery, development and therapeutics: pharmaceutical industry perspective

Data protection and connected care: provider and regulator perspectives

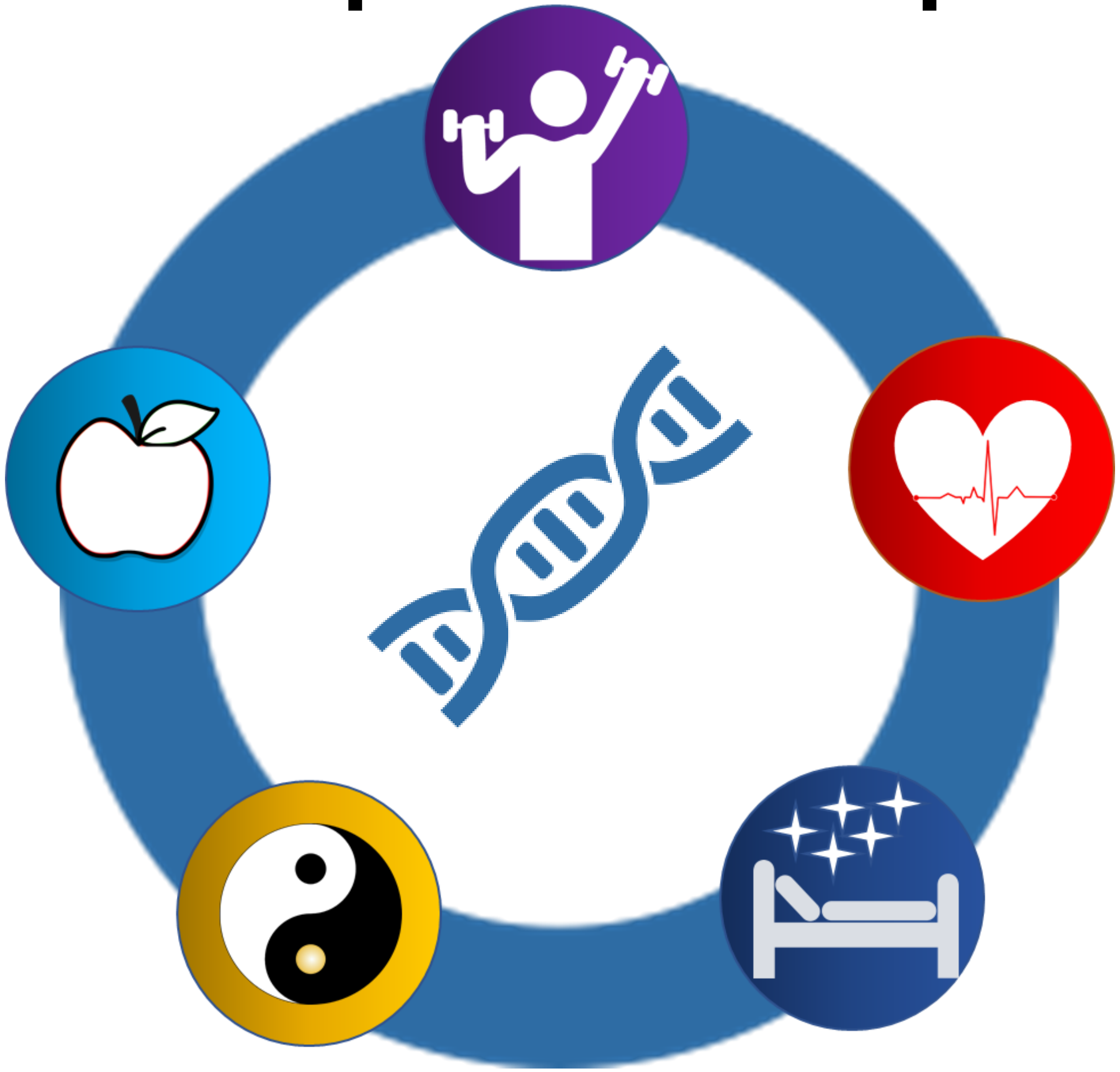
Part 2: Critical evaluation of data-driven healthcare

Traditional statistical approaches to healthcare

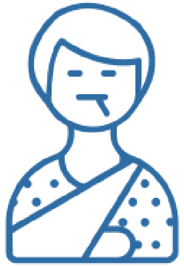
- Principles of study design
- Types of study design
- Causal modelling

Current applications of ML in the healthcare domain

Machine Learning has the Potential to Disrupt and Impact Healthcare

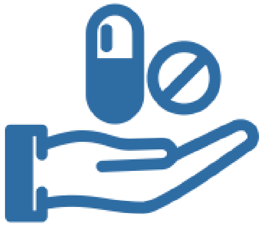


The Stakeholders in Healthcare



Patient/Person

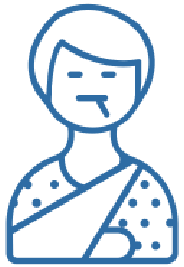
Population



Pharmaceuticals

Providers

The Person at the Centre of Healthcare



Patient/Person

ML has the capacity to transform healthcare

- Understanding physiological changes over time
- Forecasting of progression or onset of disease
- Personalising treatment strategies

Population Data-driven Healthcare

Population

What we as a society do collectively to assure the conditions in which people can be healthy

Elucidates average effects and deviations from average effects

Policy recommendations

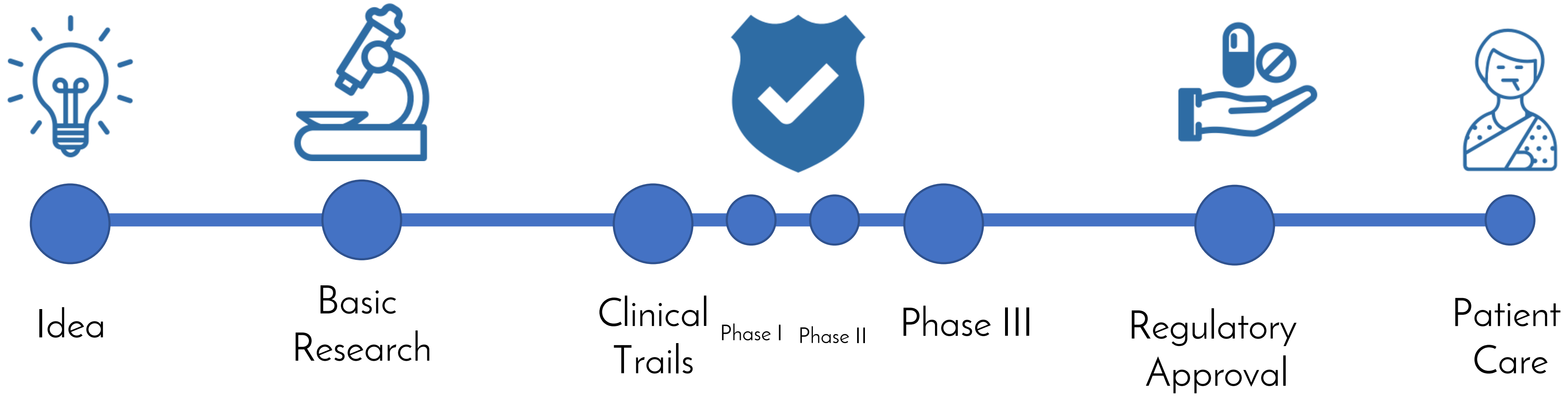
Health education

Outreach

Research for disease detection and injury prevention

Reduce healthcare inequalities

The Pharmaceutical Perspective: Drug Discovery and Therapeutics



General Data Protection Regulation

Enhance protection of personal data

Significant impact for organisations and how they manage data with some potentially very large penalties for violations – 4% of global revenues

Impacts the storage, processing, access, transfer, and disclosure of an individual's data records

These protections apply to any organisation (anywhere in the world) that processes the personal data of EU data subjects

Data Protection and Connected Care: The Provider and Regulator Perspective

Providers

FDA
U.S Food and Drug Administration

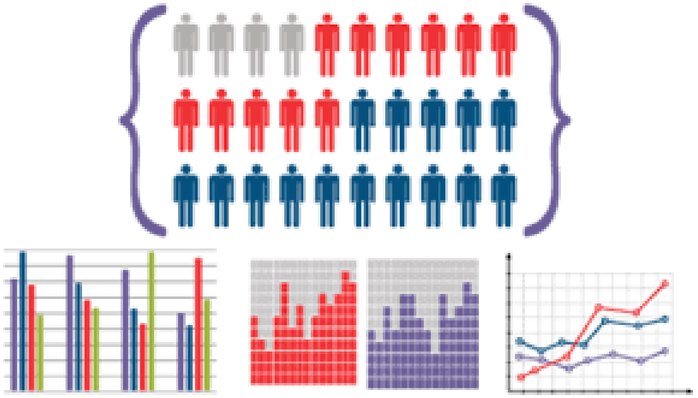
MHRA
Medicines and Healthcare products Regulatory Agency

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

AN EVALUATION OF DATA-DRIVEN HEALTH

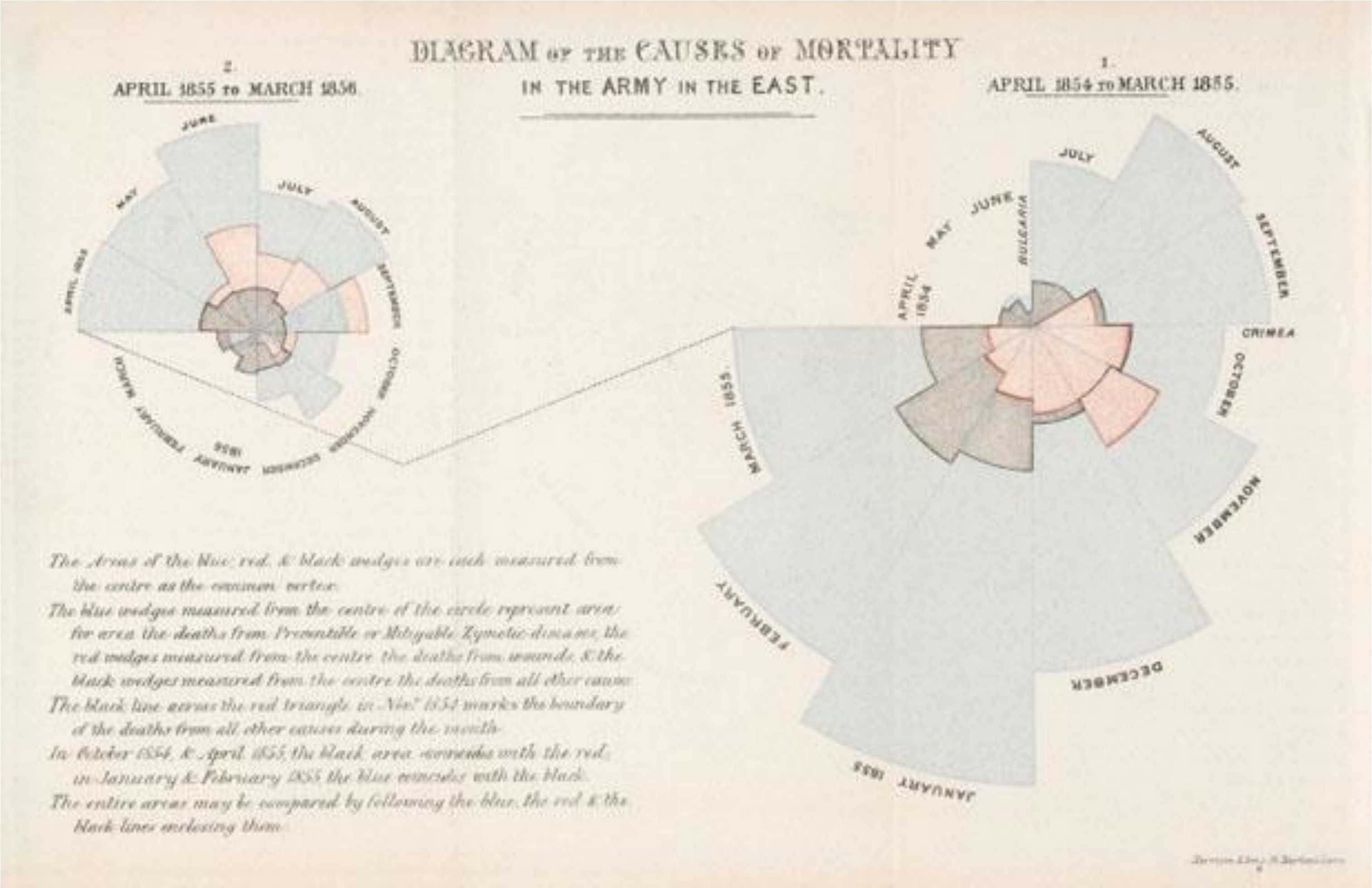
Biostatistical and Epidemiological Principles

The Beginnings of Data-Driven Health



The study of the distribution and determinants of health related states or events in specific populations & the applications of this study to the control of health problems

The Beginnings of Data-Driven Health

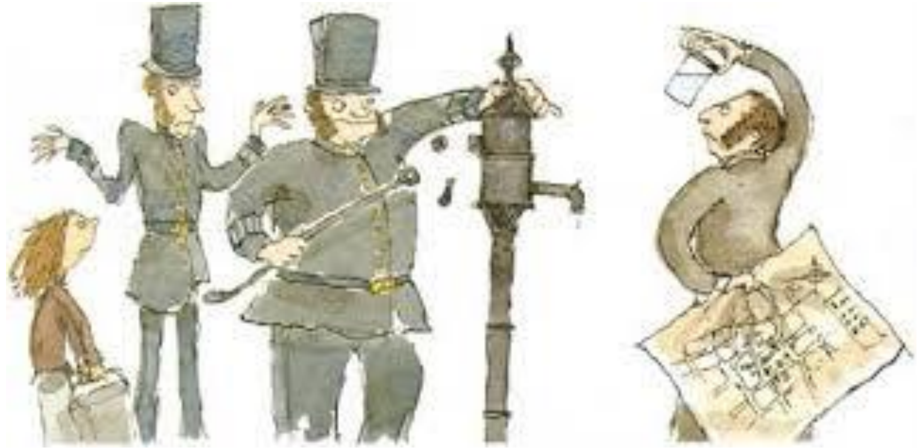


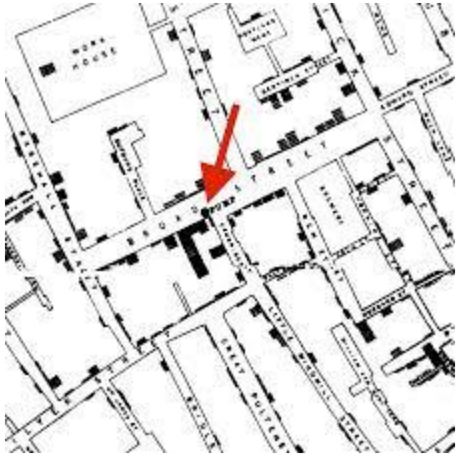
Florence Nightingale (1820 – 1910)

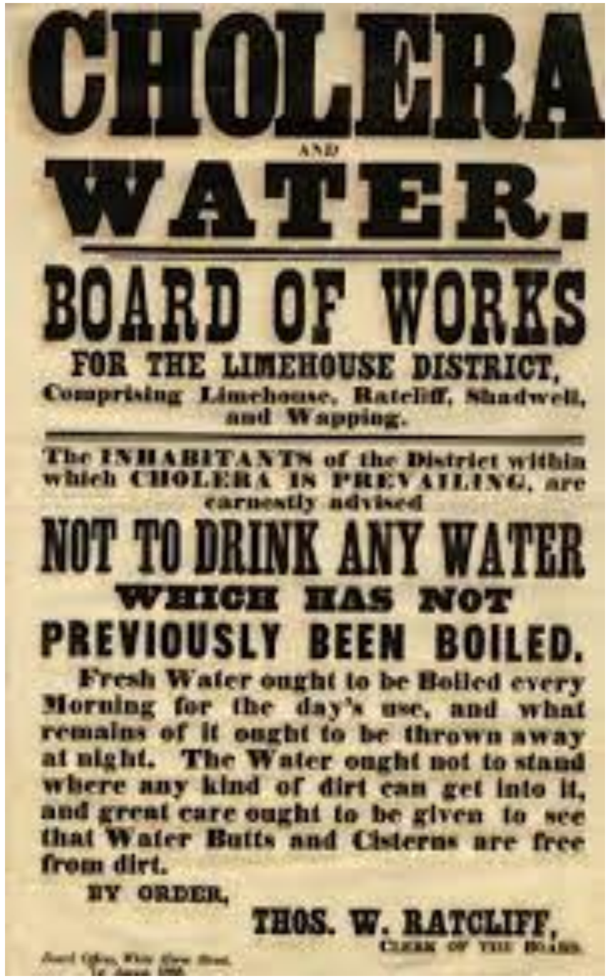
Data visualisation: death toll of the Crimean War

Army data: 16,000/18,000 deaths not due to battle wounds, but to preventable diseases, spread by poor sanitation

The Beginnings of Data-Driven Health







CHOLERA
AND
WATER.

BOARD OF WORKS
FOR THE LIMEHOUSE DISTRICT,
Comprising Limehouse, Ratcliff, Shadwell, and Wapping.

The INHABITANTS of the District within which CHOLERA IS PREVAILING, are earnestly advised

NOT TO DRINK ANY WATER WHICH HAS NOT PREVIOUSLY BEEN BOILED.

Fresh Water ought to be Boiled every Morning for the day's use, and what remains of it ought to be thrown away at night. The Water ought not to stand where any kind of dirt can get into it, and great care ought to be given to see that Water Butts and Cisterns are free from dirt.

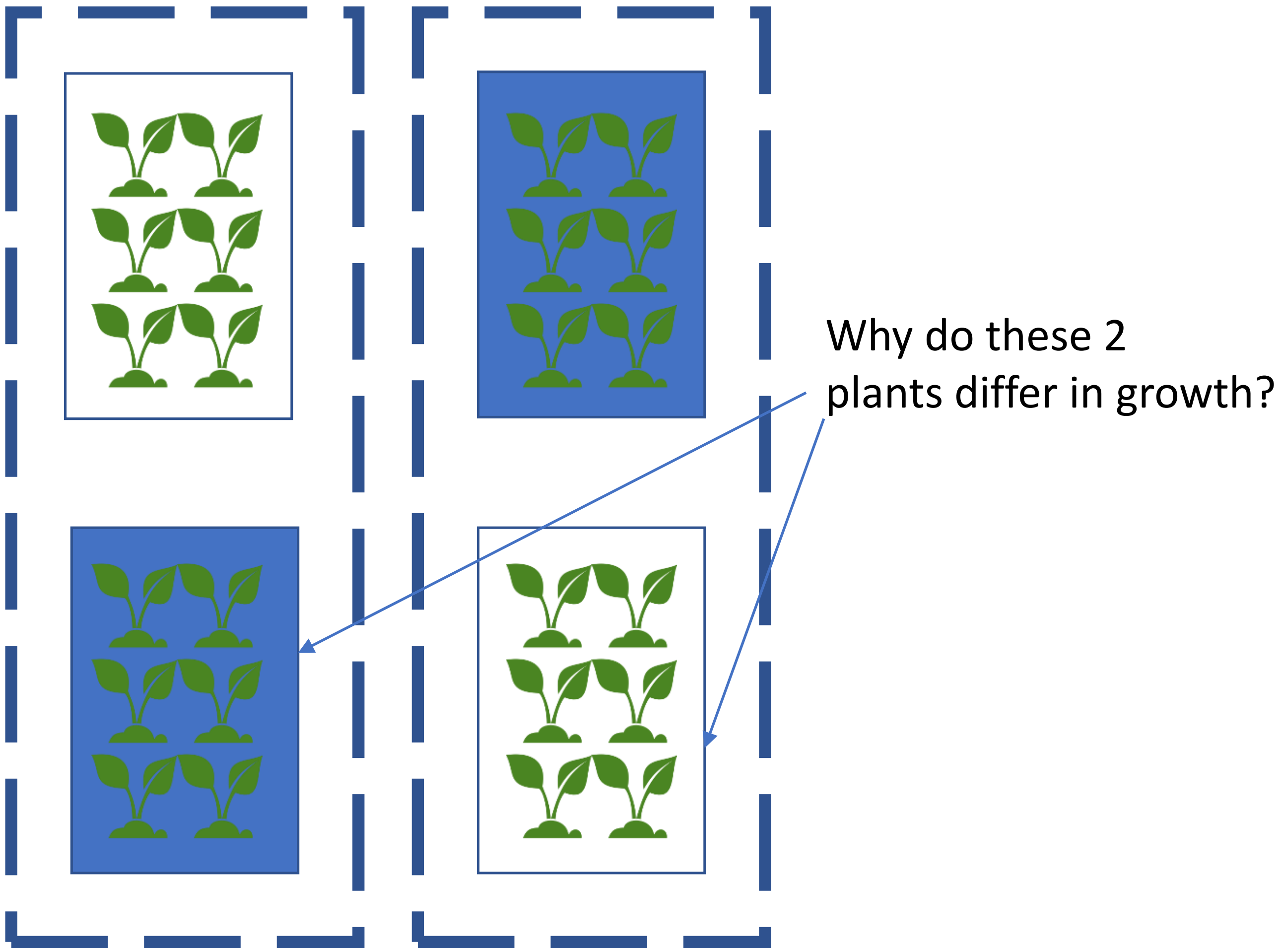
BY ORDER,
THOS. W. RATCLIFF,
CLERK OF THE BOARD.

Contextual phenomena: cholera incidence

Ecological design: compare cholera rates by region

Cohort design: compare cholera rates in exposed and non-exposed individuals

R.A. Fisher and the Principles of Experimental Design



1. Randomisation: Unbiased allocation of treatments to different experimental plot

2. Replication: repetition of the treatment to more than one experimental plot

3. Error control: Measure for reducing the error of variance

Principles of Study Design

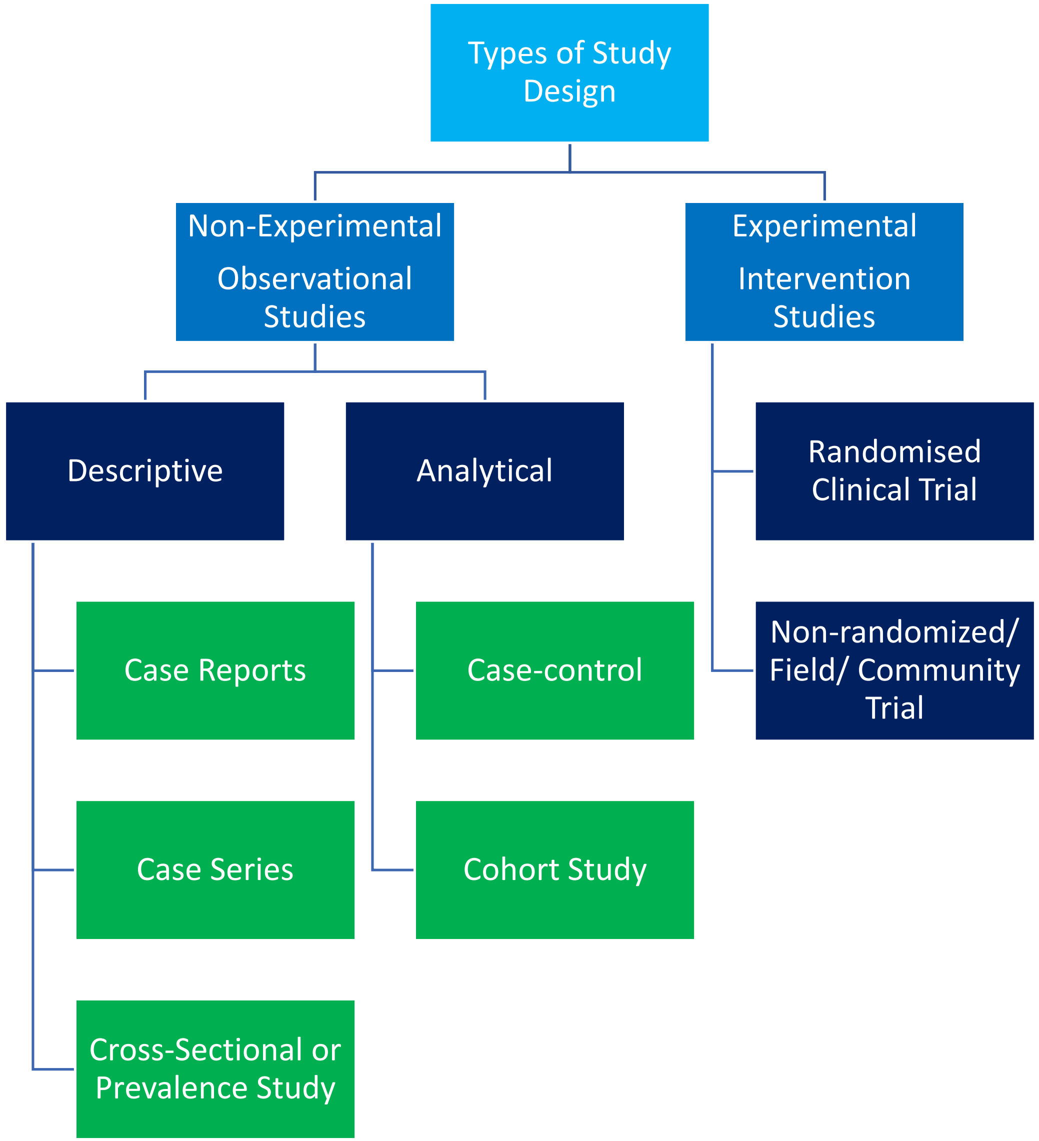
Need to set up a study to answer a research question

Design most important aspect of a study and perhaps the most neglected

The study design should match research question

- So that we don't end up collecting useless data or the principle outcome ends up not being recorded

No matter how good an algorithm is, if the study design is inadequate (garbage in) for answering the research question, we'll get garbage out


Types of Study Design
Non-Experimental
Observational Studies
Experimental
Intervention Studies
Descriptive
Analytical
Randomised Clinical Trial
Case Reports
Case-control
Non-randomized/ Field/ Community Trial
Case Series
Cohort Study
Cross-Sectional or Prevalence Study

Important Concept: Randomisation

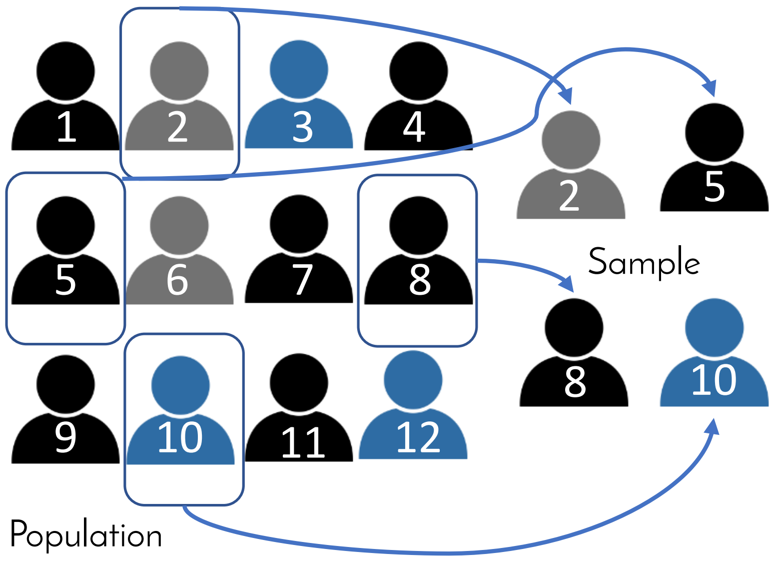
Definition: The process by which allocation of subjects to treatment groups is done by chance, without the ability to predict who is in what group

Aims:-

- To prevent statistical bias in allocating subjects to treatment groups
- To achieve comparability between the groups
- To ensure samples representative of the general population

Methods of Randomisation

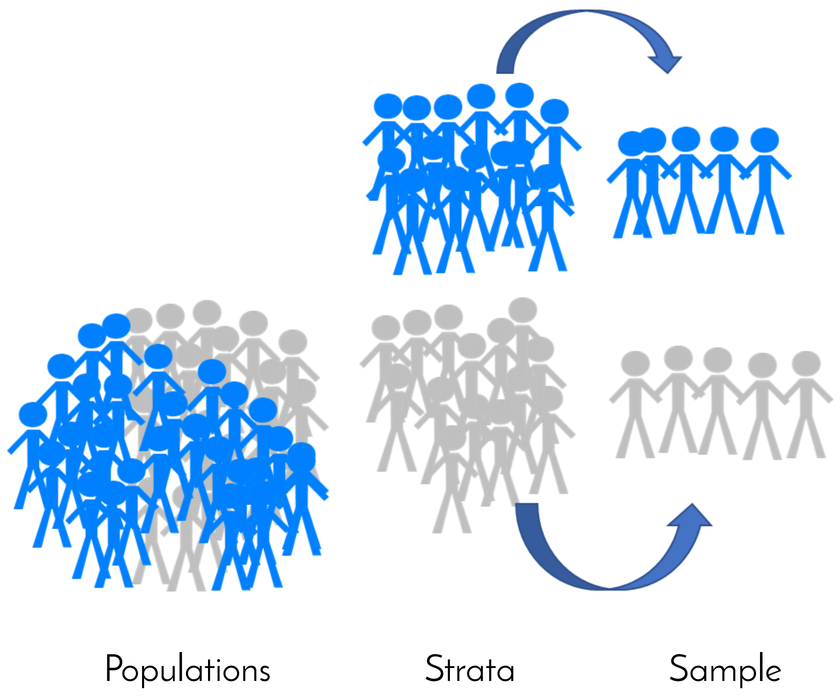
Simple Random Sampling



Permuted Block Randomisation

AABBAABB
BBBBAAA
AABAAABB

Stratified Random Sampling



Sample Size and Power Calculations

	No disease	Disease
No disease (D = 0)	☺ Specificity	X Type I error (False +) **α**
Disease (D = 1)	X Type II error (False -) **β**	☺ Power 1 - **β**; Sensitivity

Power is the probability that a test of significance will pick up on an effect that is present

Increases with

- sample size
- effect size
- type I error

$$Power \propto \frac{Sample\ size\ (n)}{Effect\ size\ (\Delta), Alpha(\alpha)}$$

The Challenge of Missing Data

Missing data is a common problem in healthcare data and can produce biased parameter estimates

Reasons for missingness may be informative for estimating model parameters

Bayesian models: coherent approach to incorporating uncertainty by assigning prior distributions

Mason, Alexina, Nicky Best, Sylvia Richardson, and IAN PLEWIS. "Strategy for modelling non-random missing data mechanisms in observational studies using Bayesian methods." Journal of Official Statistics (2010)

Missing Data

Missing Completely At Random (MCAR)

The probability of data being missing does not depend on the observed or unobserved data

e.g. $\text{logit}(p_{it}) = \theta_0$

Missing At Random (MAR)

The probability of data being missing does not depend on the unobserved data, conditional on the observed data

e.g. Children with missing wheeze data have better lung function

e.g. $\text{logit}(p_{it}) = \theta_0 + \theta_1 t_i$ or $\text{logit}(p_{it}) = \theta_0 + \theta_2 y_0$

Missing Not At Random (MNAR)

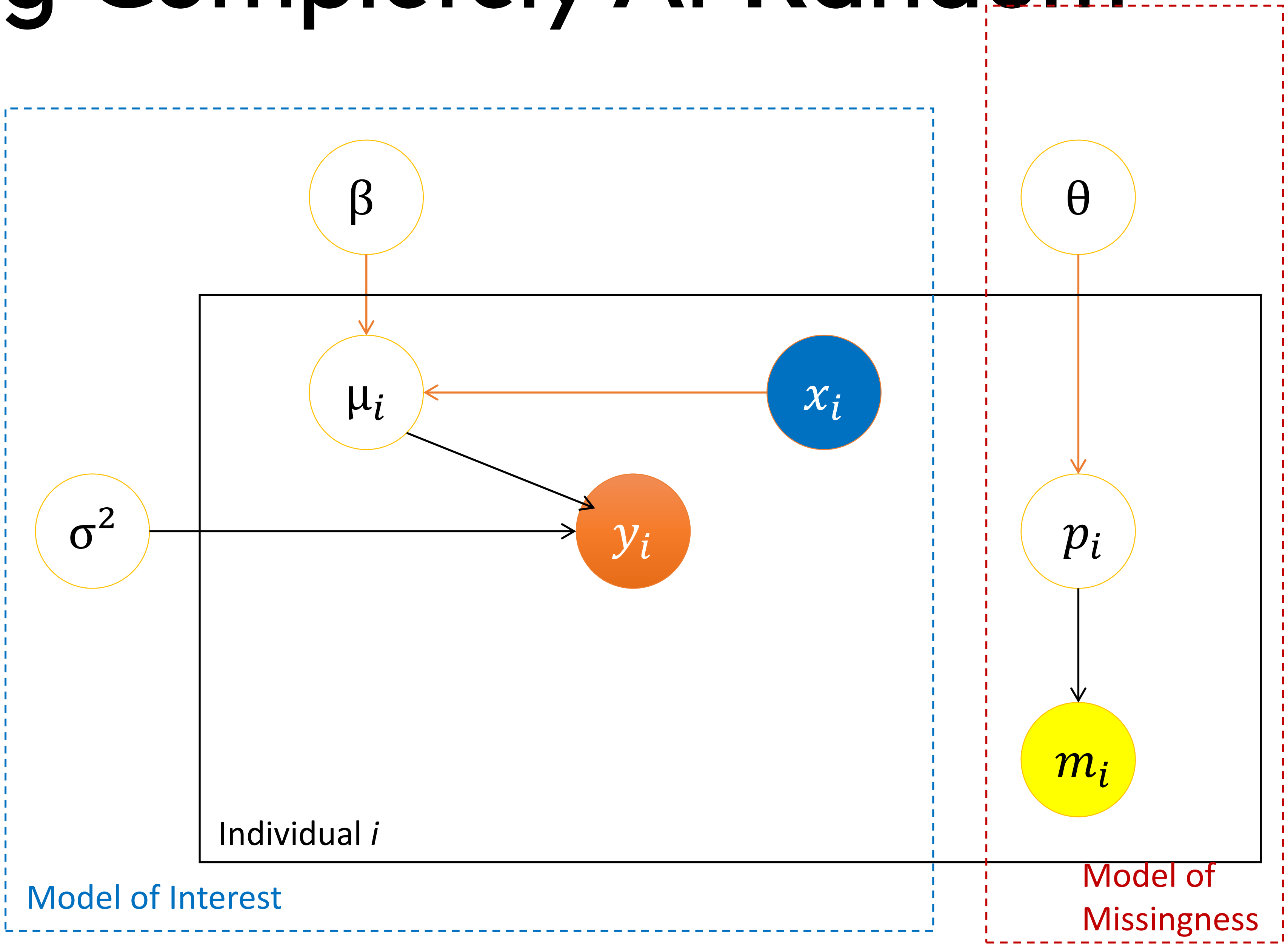
The probability of data being missing does depend on the unobserved data, conditional on the observed data.

e.g. Children with missing lung function have better lung function

e.g. $\text{logit}(p_{it}) = \theta_0 + \theta_3 y_{it}$

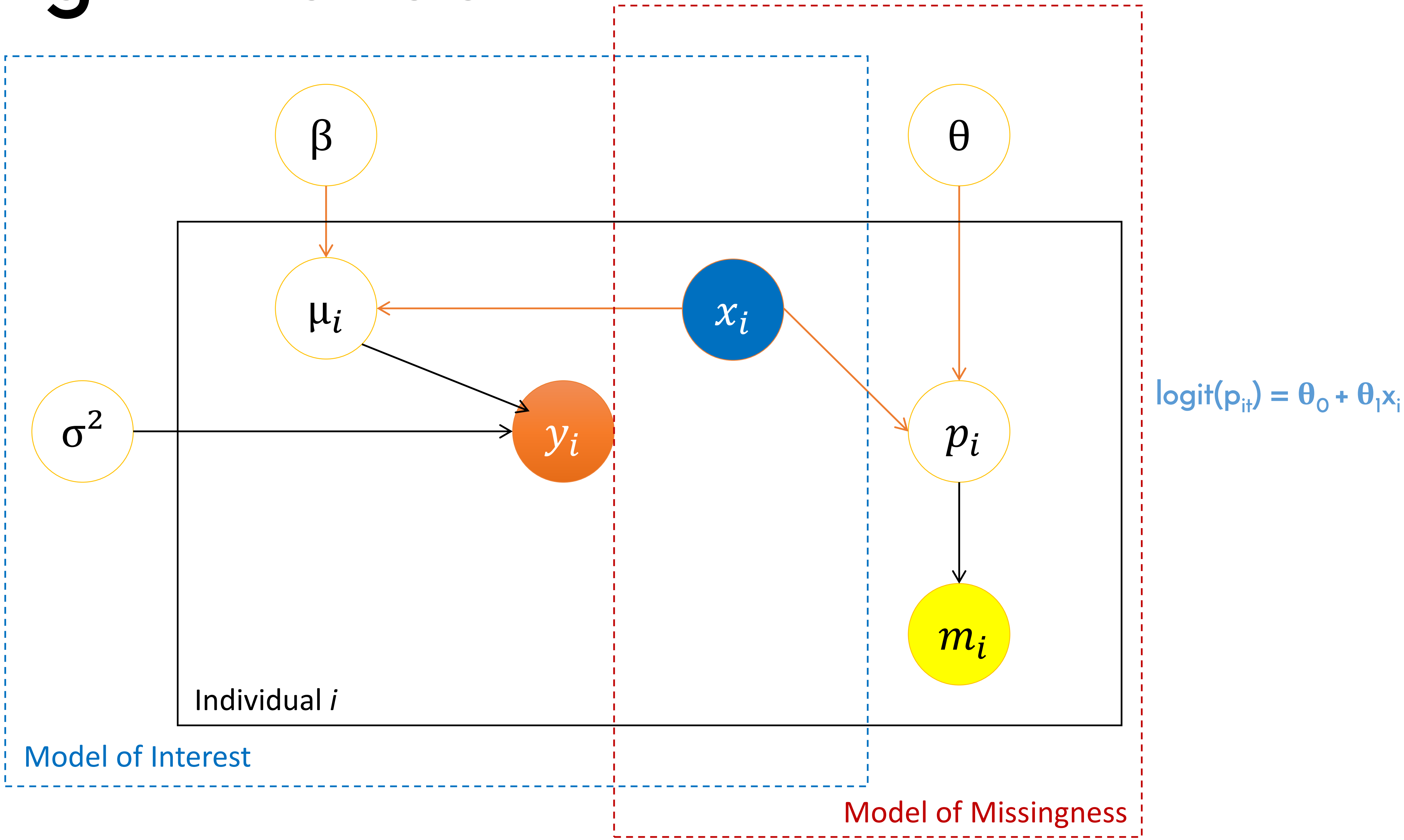
Alexina Mason. "Bayesian methods for modelling non-random missing data mechanisms in longitudinal studies" PhD Thesis (2009)

Missing Completely At Random

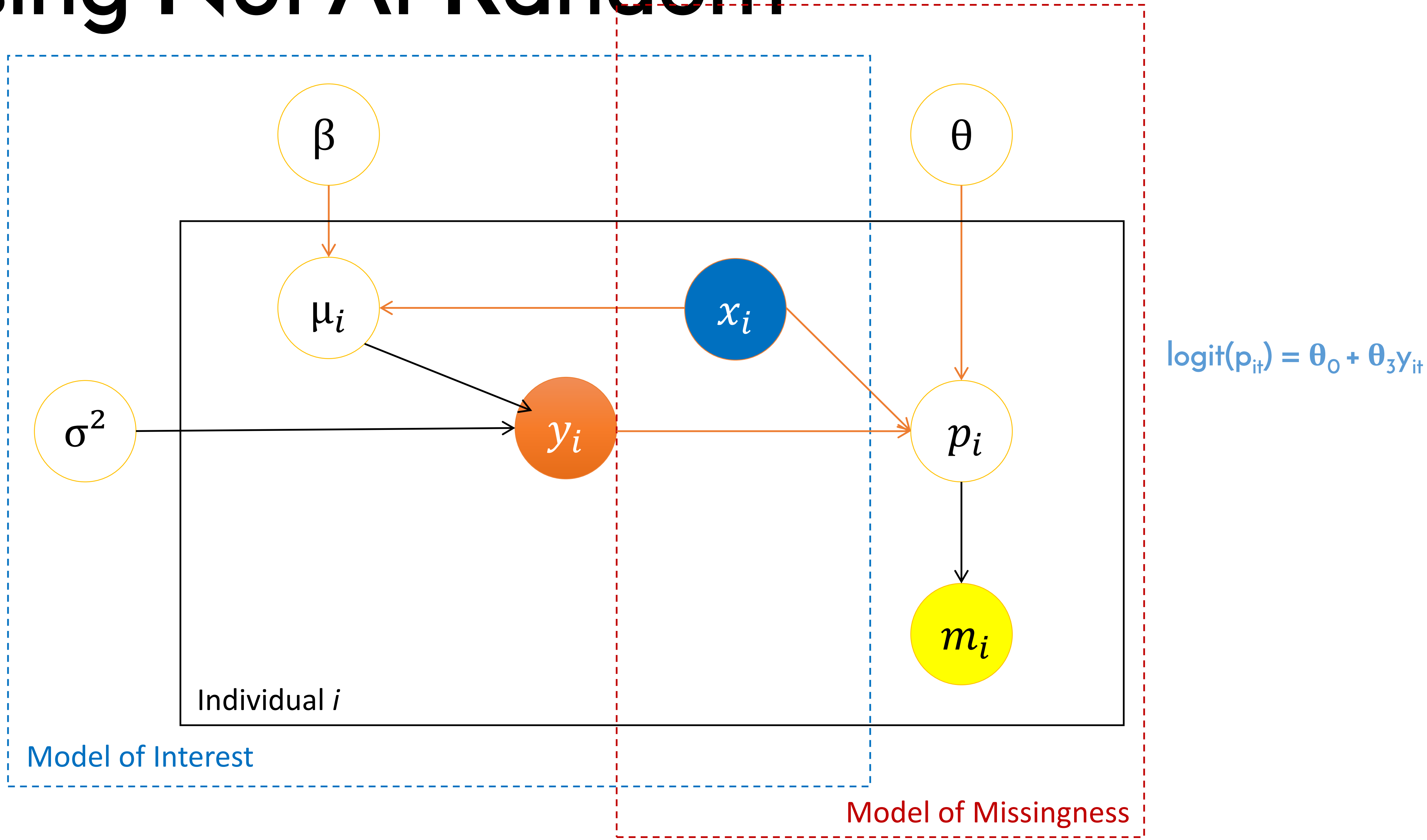


$\mathrm{logit}(p_{it}) = \theta_0$

Missing At Random



Missing Not At Random



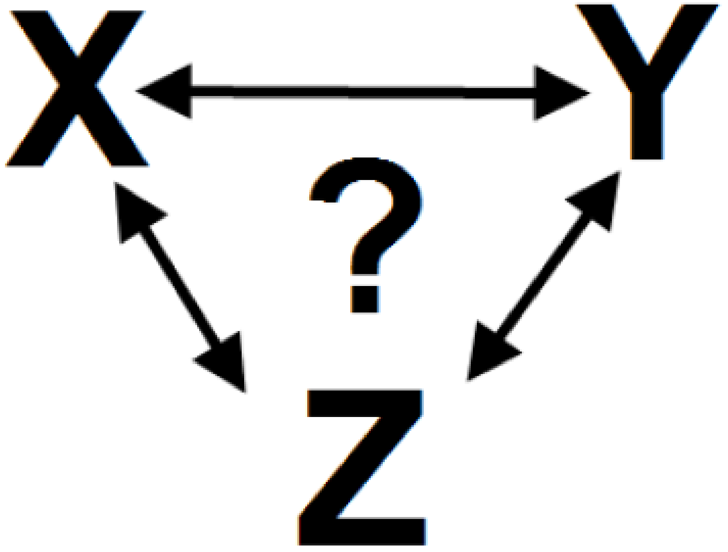
Causal Reasoning

The questions that motivate most studies in the health, social and behavioral sciences are not associational but causal in nature.

Before an association is assessed for the possibility that it is causal, other explanations such as chance, bias and confounding have to be excluded

Require some knowledge of the data-generating process - cannot be computed from the data alone, nor from distributions governing data

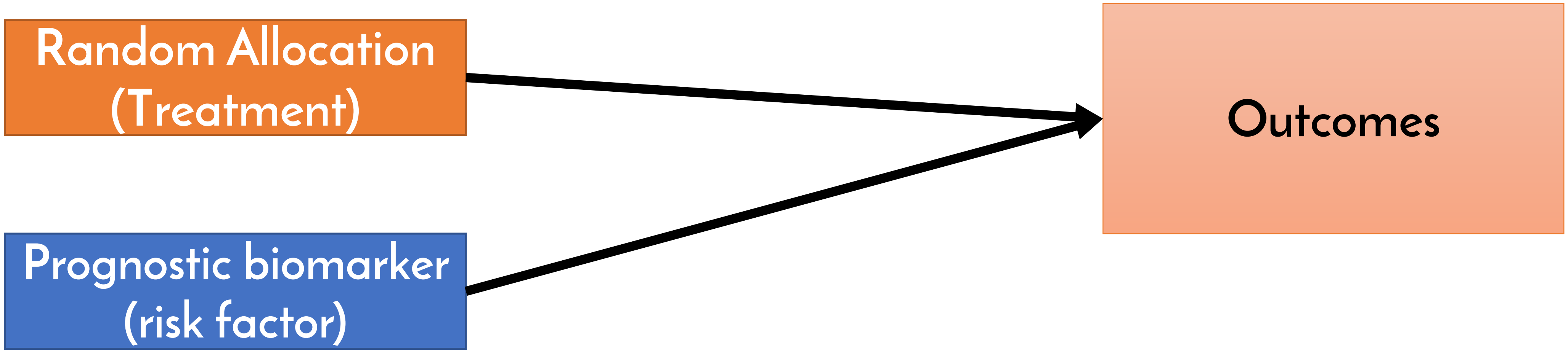
Aim: to infer dynamics of beliefs under changing conditions, for example, changes induced by treatments or external interventions.



Pearl, Judea. "Causal inference in statistics: An overview." *Statistics surveys* 3 (2009): 96-146.

Prognostic Biomarker (Risk Factor)

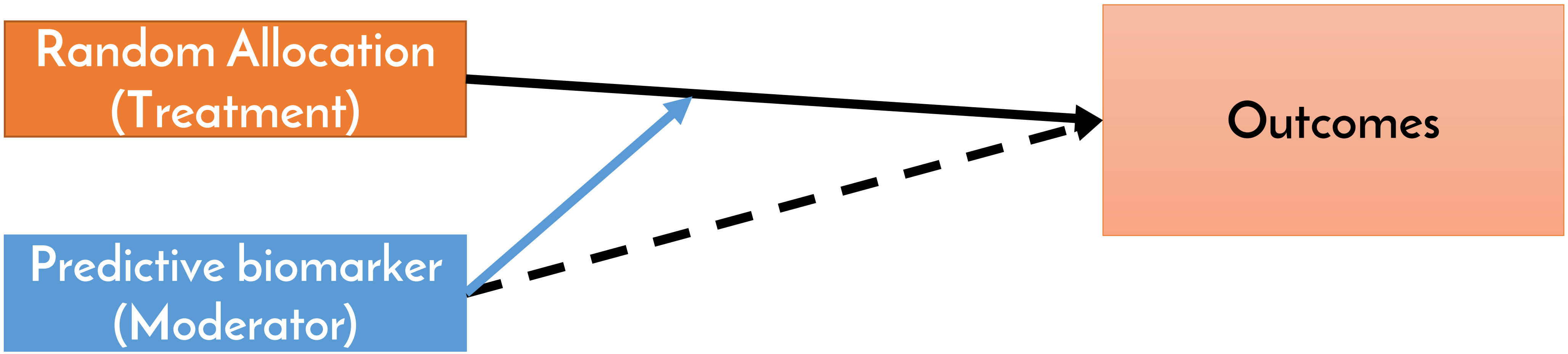
A biological measurement made before treatment to indicate long-term outcome for patients either untreated or receiving standard outcome



Dunn, Graham, Richard Emsley, Hanhua Liu, and Sabine Landau. "Integrating biomarker information within trials to evaluate treatment mechanisms and efficacy for personalised medicine." *Clinical Trials* 10, no. 5 (2013): 709-719.

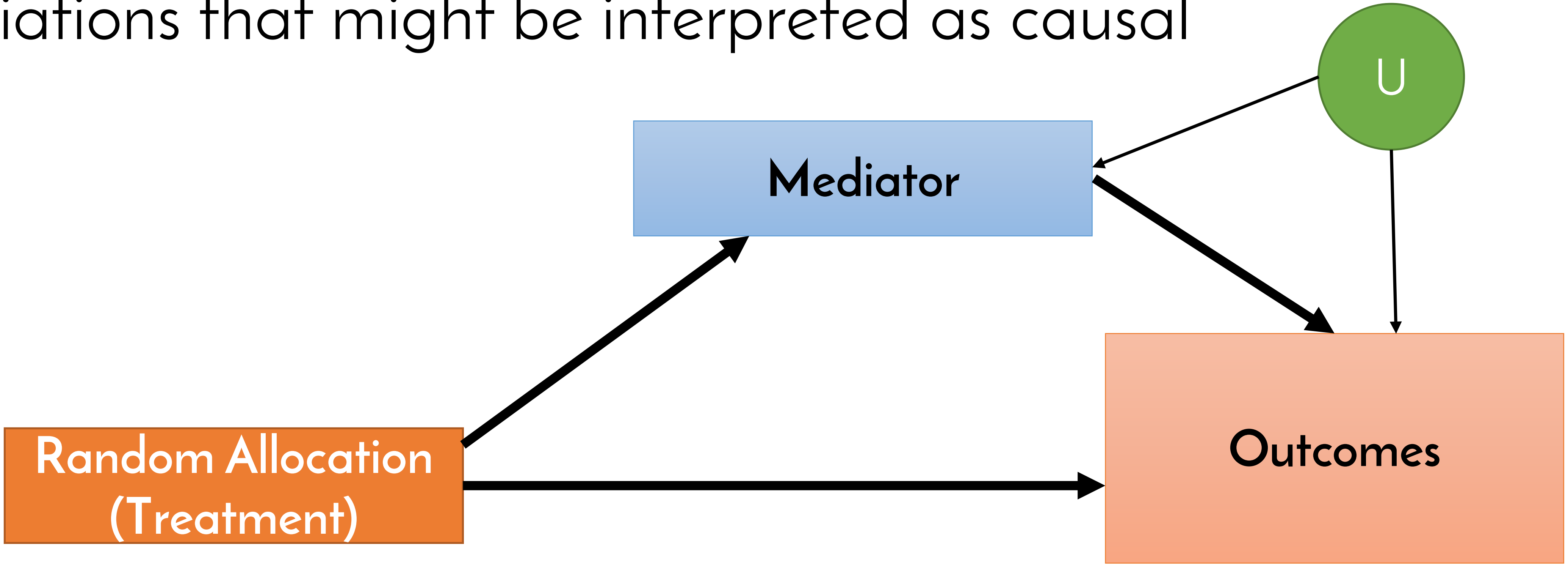
Predictive Biomarker (Moderator)

A variable that **changes the impact** of treatment on the outcome. A biological measurement made before treatment to identify patients likely or unlikely to benefit from a particular treatment

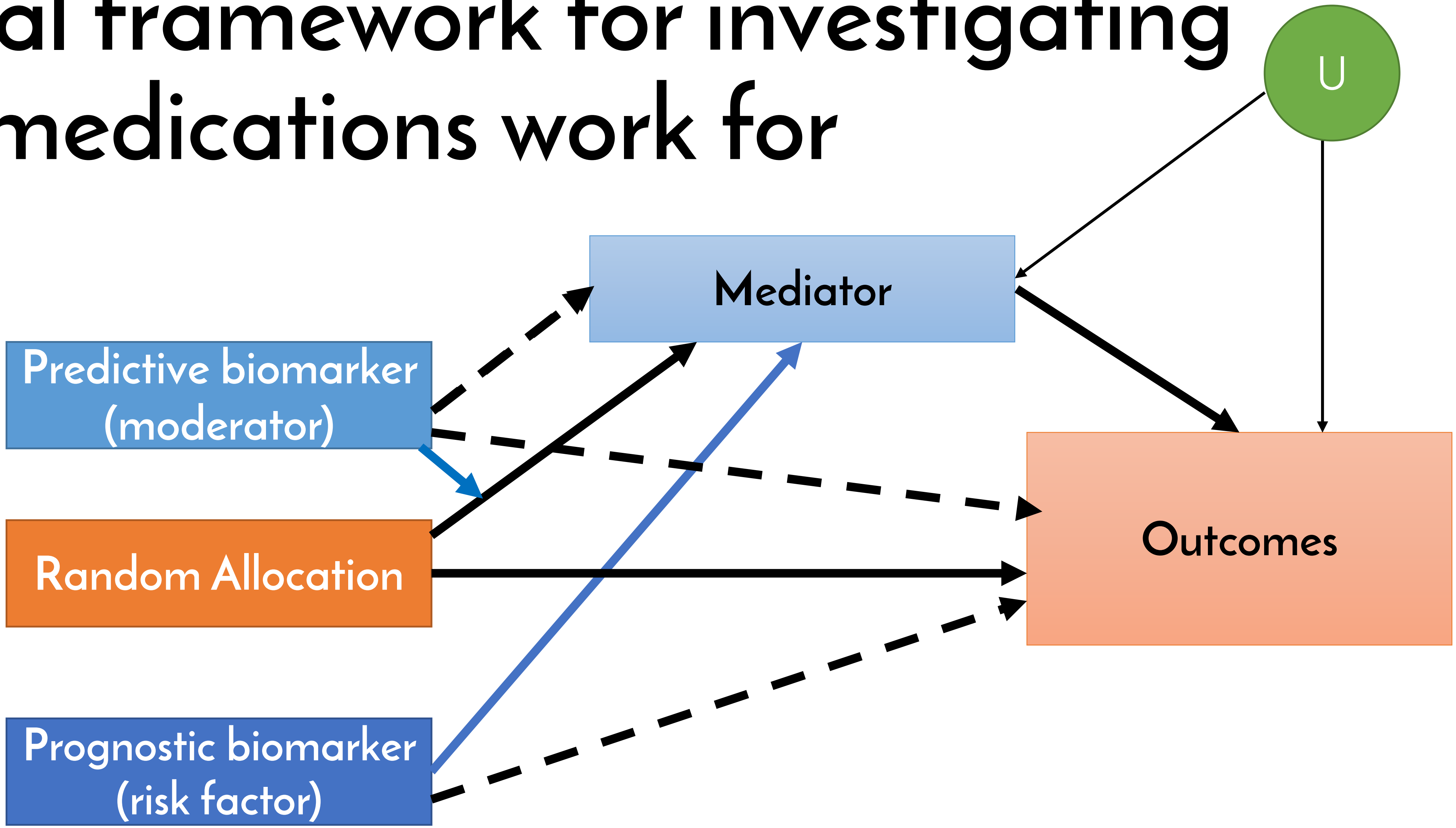


Mediator

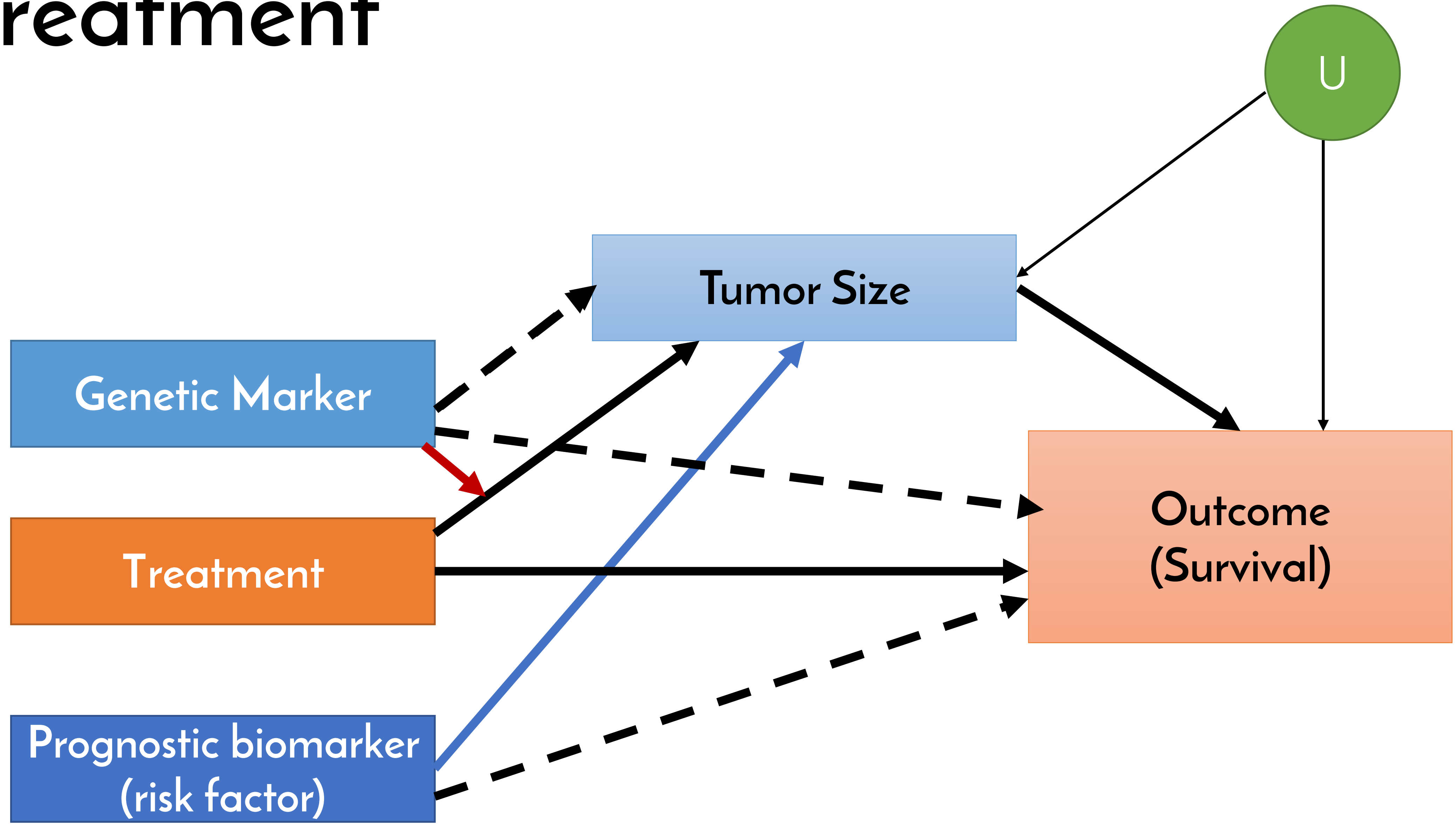
A **mechanism** by which one variable affects another variable. Omitted common causes (hidden confounding) should always be considered as a possible explanation for associations that might be interpreted as causal



Efficacy and mechanism evaluation: Causal framework for investigating who medications work for



Example: Personalisation of Cancer Treatment



Bradford-Hill Principles of Causality

Plausibility
Does causation make sense

Consistency
Cause associated with disease in different population and studies

Temporality
Cause precedes disease

Strength
Cause strongly associated with disease

Specificity
Does the cause lead to a specific effect

Dose-Response
Greater exposure to cause, higher the risk of disease

Machine Learning for Healthcare in Context

Health data complexity requires adequately complex methodologies and algorithms

Methods don't scale, need more advanced techniques and thinking about other techniques developed outside the traditional stats community

Need for scale and speed

CURRENT WORK IN ML IN THE HEALTHCARE DOMAIN

Data complexity requires adequately complex algorithms

Sparsity in Health Data

Major challenge for truly generalizable and scalable AI in healthcare is maximizing information utility for public health impact when that information (observational or clinical-context data) is sparse

- Missing data
- Inadequately sampled data
- Data that does not represent the diversity of a population

Generalisability: Training datasets that are representative of the diversity of the population as well as the heterogeneity of health conditions.

Transfer learning: potential to

- Maximise utility of available data
- Improve model's ability to generalise

Transfer Learning for Data Sparsity

Good quality healthcare data is expensive and very often sparse

Aim: Maximizing information by using multiple data sources

Challenge: Feature mismatch: features in different datasets may vary

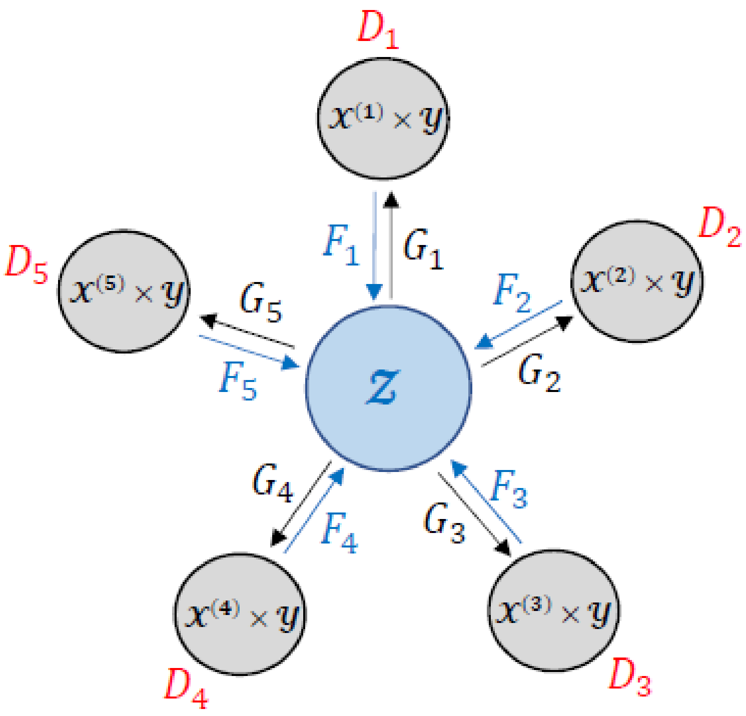
Challenge: Distribution Mismatch: differing patient populations across different hospitals

GAN architectures to efficiently enlarge the dataset

Better predictive models than if we simply used the target dataset

Jinsung Yoon, James Jordon and Mihaela van der Schaar.
"RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks." arXiv preprint arXiv:1802.06403 (2018).

RadialGAN Transfer Learning for Data Sparsity



Z: Latent space
$X^{(i)}$ x Y: i^{th} domain
G_i, F_i, D_i: Decoders, Encoders and Discriminator of the i^{th} domain

The i^{th} domain is translated to the j^{th} domain via Z using F_i and G_j

Jinsung Yoon, James Jordon and Mihaela van der Schaar.
"RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks." arXiv preprint arXiv:1802.06403 (2018).

npj | Digital Medicine www.nature.com/npjdigitalmed

ARTICLE OPEN

The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants

Elad Yom-Tov[1], Jinia Shembekar[2], Sarah Barclay[2] and Peter Muennig[3]

As public health advertisements move online, it becomes possible to run inexpensive randomized-controlled trials (RCTs) thereof. Here we report the results of an online RCT to improve food choices and integrate exercise into daily activities of internet users. People searching for pre-specified terms were randomized to receive one of several professionally developed campaign advertisements or the "status quo" (ads that would otherwise have been served). For 1-month pre-intervention and post-intervention, their searches for health-promoting goods or services were recorded. Our results show that 48% of people who were exposed to the ads made future searches for weight loss information, compared with 32% of those in the control group—a 50% increase. The advertisements varied in efficacy. However, the effectiveness of the advertisements may be greatly improved by targeting individuals based on their lifestyle preferences and/or sociodemographic characteristics, which together explain 49% of the variation in response to the ads. These results demonstrate that online advertisements hold promise as a mechanism for changing population health behaviors. They also provide researchers powerful ways to measure and improve the effectiveness of online public health interventions. Finally, we show that corporations that use these sophisticated tools to promote unhealthy products can potentially be outbid and outmaneuvered.

npj Digital Medicine (2018)1:24 ; doi:10.1038/s41746-018-0031-7

INTRODUCTION

Hundreds of millions of dollars are spent on traditional public health advertisements annually.[1–7] In theory, public health advertising can save money and lives by encouraging behaviors that prevent disease before it happens.[8] While the objective of advertising investments (e.g., encouraging people to quit smoking) differs from those of private advertisers (encouraging people to purchase a good or service), the central idea is the same: to change behaviors.

Before online advertising, it was only possible to empirically test public health campaigns by randomizing small numbers of participants and to examine a few outcome measures.[1,2] This makes it difficult to test to whom different forms of advertisement are best targeted.[3–6]

Humans vary greatly with respect to both their biology and their beliefs. Medical researchers use predictive analytics to mine databases of genetic information in order to target treatments to individuals who are more likely to respond to them. Similarly, private advertisers use predictive analytics to mine multiple sources of sociodemographic and behavioral data to better target individual consumers with the goal of changing their behavior. However, precision public health interventions have largely sat on the sidelines both due to the large sums of money required for targeted advertising and due to ethical concerns.

Ethical concerns arise for a number of reasons. First, participant data are collected without informed consent.[9] Second, many in public health feel uncomfortable with the idea of manipulating individual behaviors, preferring instead to work with anonymous means to attempt to change behavior more generically.[10,11] Such concerns have largely pre-empted the use of precision public health advertising, leaving only private firms to employ these tools.

In the private sector, Google, Microsoft, Facebook, and other internet-based companies provide online services for free in exchange for the information that drives precision advertising using "big data analytics". Online ads targeted using data analytics can influence emotions and behaviors.[10,12,13]

First, advertisers can make educated guesses or small-scale tests about who might respond most to a given advertisement based on common search terms by topic. Then, advertisement can be randomized to be shown to users of search engines that search for such terms. Randomization provides a "gold standard" test of efficacy. Randomization can also provide causal information on how different sub-groups (e.g., young women) respond to an advertisement relative to others. Information on the experimental responses of different "architypes" of individuals can then re-tested with newer, more effective advertisements. This incremental approach—targeting, refining, and testing—has the power to produce online ads that affect beliefs and behaviors.

Big data companies—such as Facebook, Google, and Microsoft—conduct tens of thousands of randomized-controlled trials (RCTs) on their users every year.[14] These results are invariably kept inside these companies, but the general process for evaluating advertisement efficacy is likely similar across companies.

Search advertisements are typically presented as textual advertisements that appear on a search results page coupled with a click through link to the advertiser's site. More advanced versions include images in addition to (or instead of) the text. While it is rare that users click on ads, online advertisements have

[1]Microsoft Research Israel, 13 Shenkar st., 46875 Herzeliya, Israel; [2]J. Walter Thompson, 466 Lexington Avenue, New York, NY 10017, USA and [3]Global Research Analytics for Population Health and the Department of Health Policy and Management, Mailman School of Public Health, Columbia University, 722 West 168th St., New York, NY 10032, USA
Correspondence: Elad Yom-Tov (eladyt@microsoft.com)

Received: 28 February 2018 Revised: 30 March 2018 Accepted: 9 April 2018
Published online: 27 June 2018

Randomized control trial based on searches

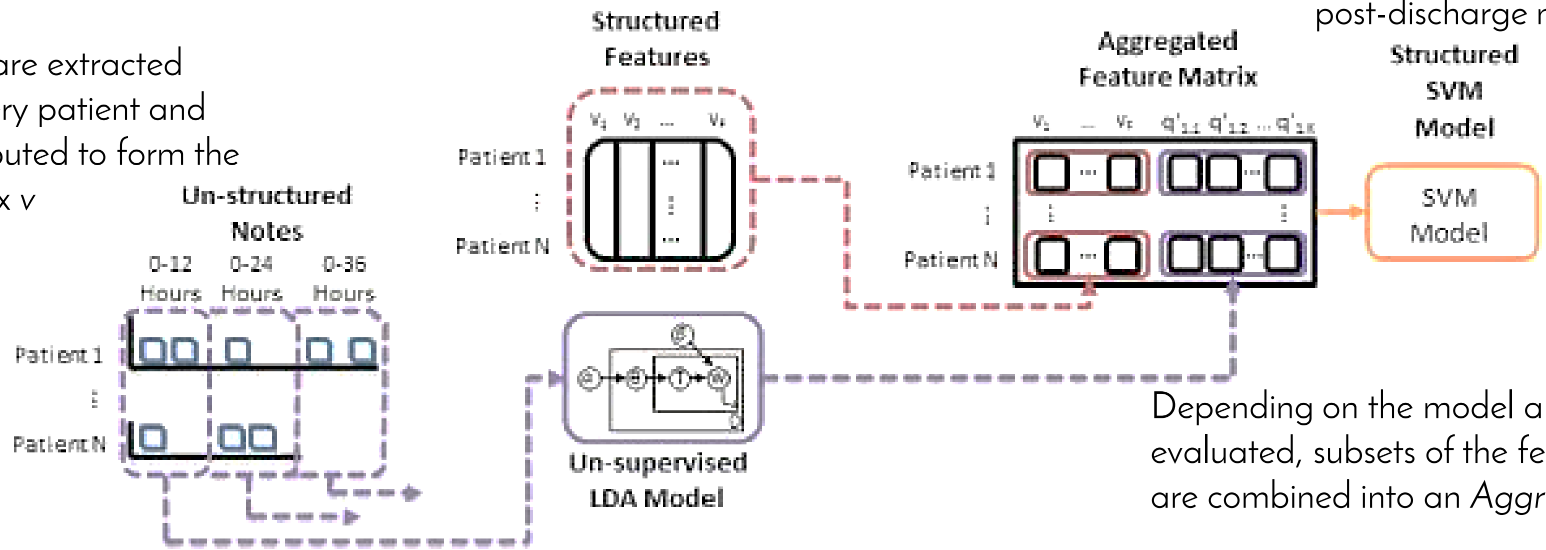
Public Health intervention: Explored the impact of ads on changing health behaviours as measured by future health promotion searches

Learning Structure from Real-World EHRs

Per-note latent topic features are aggregated in extending 12 hour windows and used to form matrix *q´* where is the overall proportion of topic *k* in time-window *m*

A linear kernel SVM is trained to create classification boundaries forthree clinical outcomes: in-hospital mortality, 30 day post-discharge mortality, and 1 year post-discharge mortality

Clinical baseline features are extracted from the database for every patient and derived features are computed to form the *Structured Features* matrix *v*



Depending on the model and time window being evaluated, subsets of the feature matrix *v* and matrix *q´* are combined into an *Aggregated Feature Matrix*

Each patient's de-identified clinical notes are used as the observed data in an LDA topic model and a total of 50 topics are inferred to create the per-note topic proportion matrix *q*

Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A. and Szolovits, P., 2014, August. Unfolding physiological state: Mortality modelling in intensive care units. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 75-84). ACM.

CONTEXTUAL EVALUATION OF PROBLEM-LED MODELLING FRAMEWORKS

Think deeply about the clinical context. Find solutions which are specific to the problem.

Good science is about merging different schools of thought for developing the bigger picture.

Data driven approach + Domain Knowledge = Problem-led approach with the patient at the centre

Problem-led vs Data-driven Health

Danielle Belgrave, John Henderson, Angela Simpson, Iain Buchan, Christopher Bishop, and Adnan Custovic.
"Disaggregating asthma: Big investigation versus big data." Journal of Allergy and Clinical Immunology 139, no. 2 (2017): 400-407..

From Information to Knowledge

1. **Team Science**: Discoveries about healthcare, **not hypothesised** a priori, have been made by experts explaining **structure** learned from **data** by algorithms tuned by those **experts**

2. Heuristic blend of **biostatistics** and **machine-learning** for principled problem-led healthcare research

3. An ML approach to extracting knowledge from information in healthcare requires persistent integration of
   Data
   Methods
   Expertise

Problem-Led Patient-Centred Research

THANK YOU

Microsoft

Imperial College London

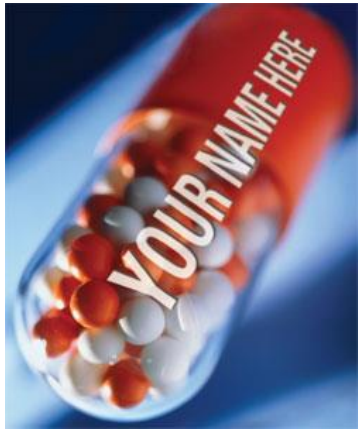
THE UNIVERSITY OF SYDNEY

ML STRATEGIES FOR HEALTHCARE PERSONALISATION

Konstantina Palla

Microsoft

WHAT IS PERSONALISED HEALTHCARE



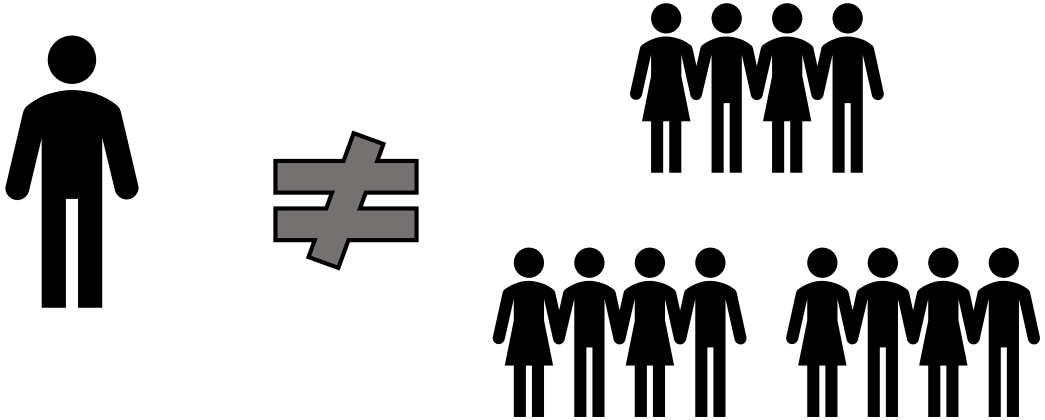
- Traditionally -> *personalised medicine*

"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

Patient



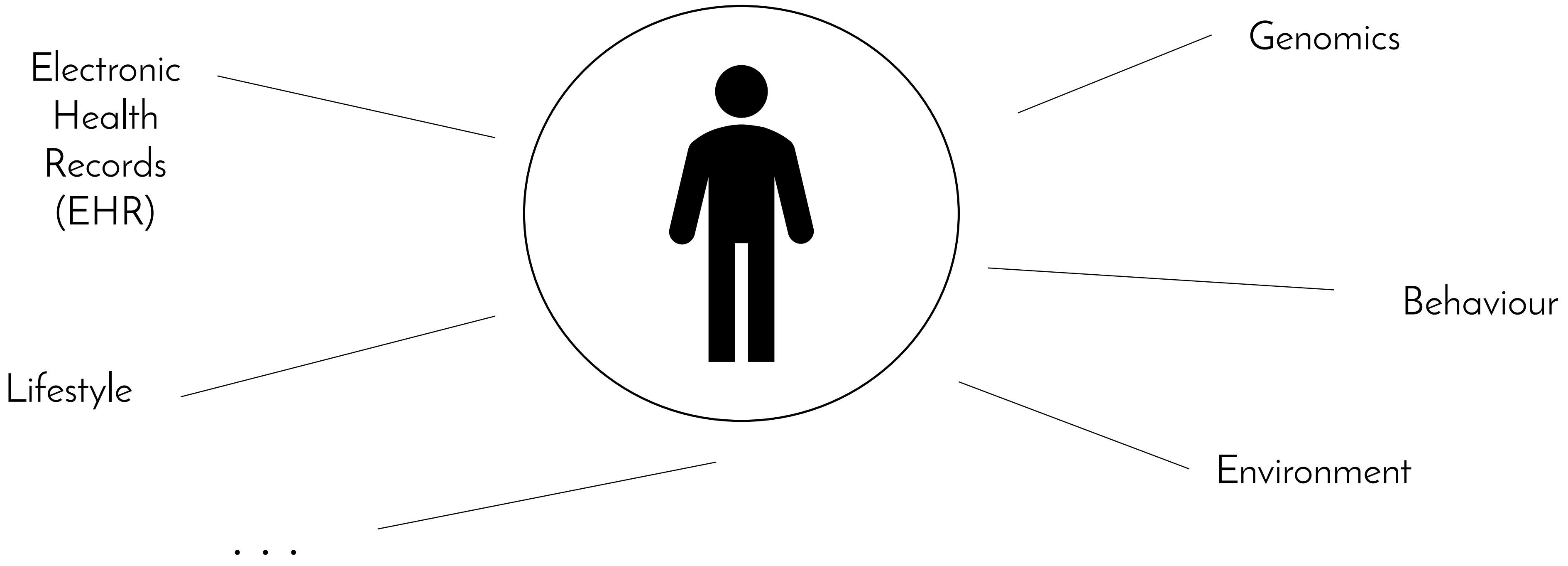
Factors of disease heterogeneity:

Genomics
Behaviour
Prior exposures
Comorbidities
Etc.

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

WHAT IS PERSONALISED HEALTHCARE

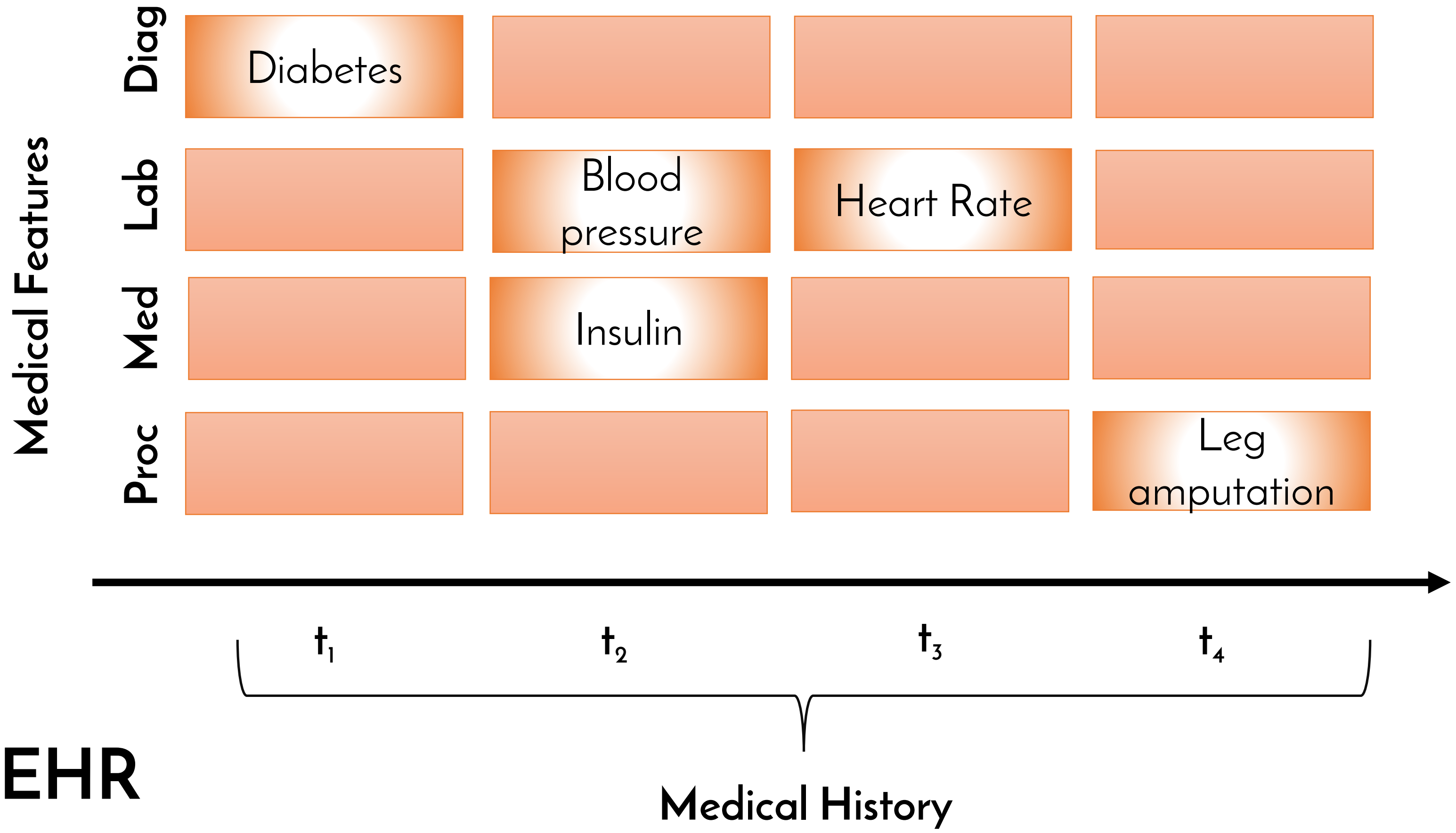
Person in the centre. Person as unique individual.

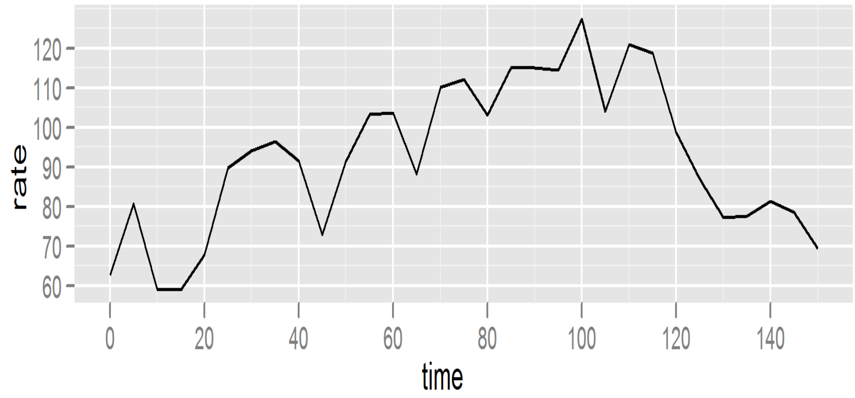


Provision of Prognosis, Diagnosis, Treatment **tailored** to the individual

PERSONALISED HEALTHCARE – HOW CAN ML HELP?

ML can transform data into actionable information

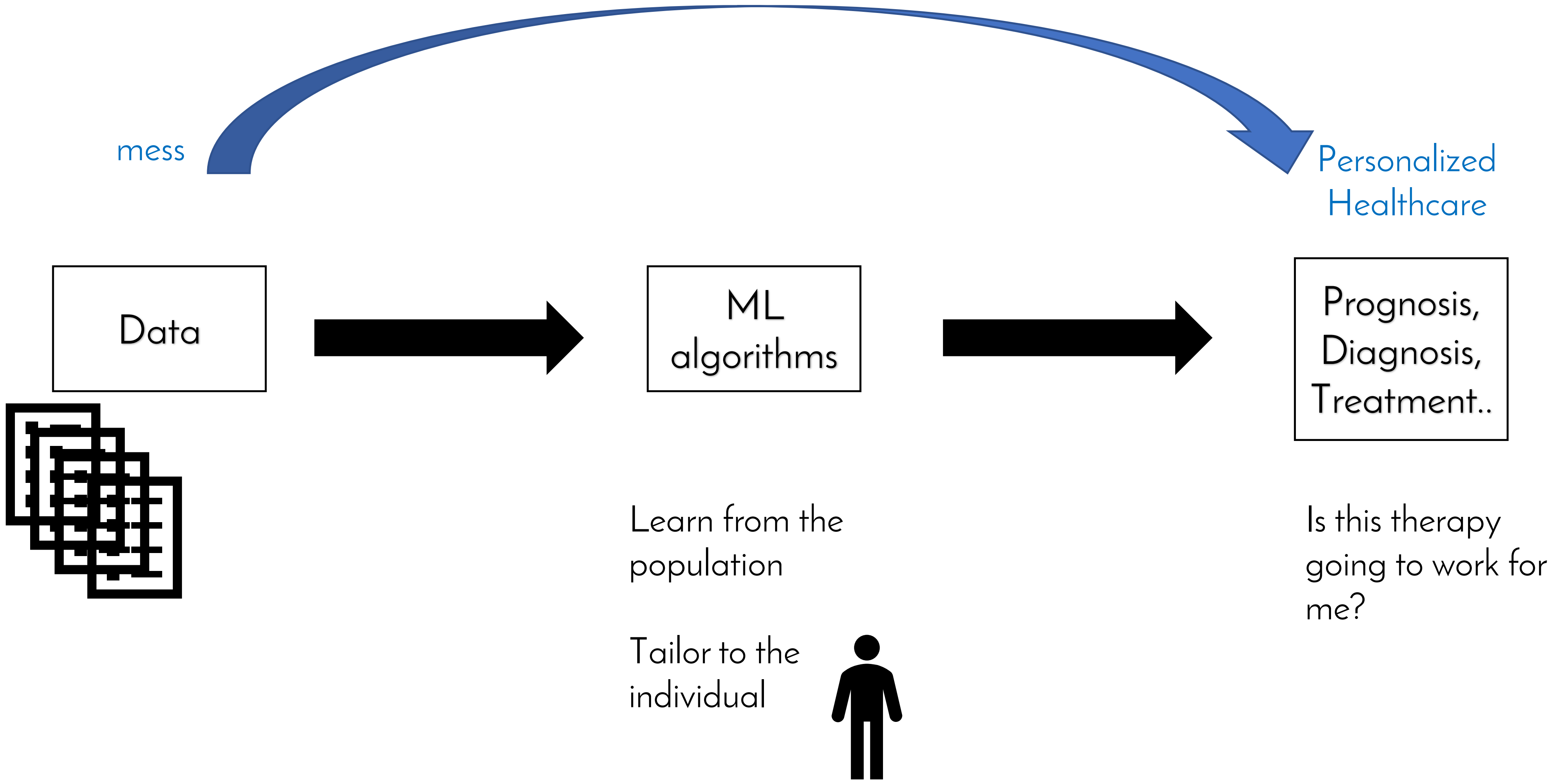




How can we extract useful knowledge?

Inspired by [**Lee et al., 2017**]

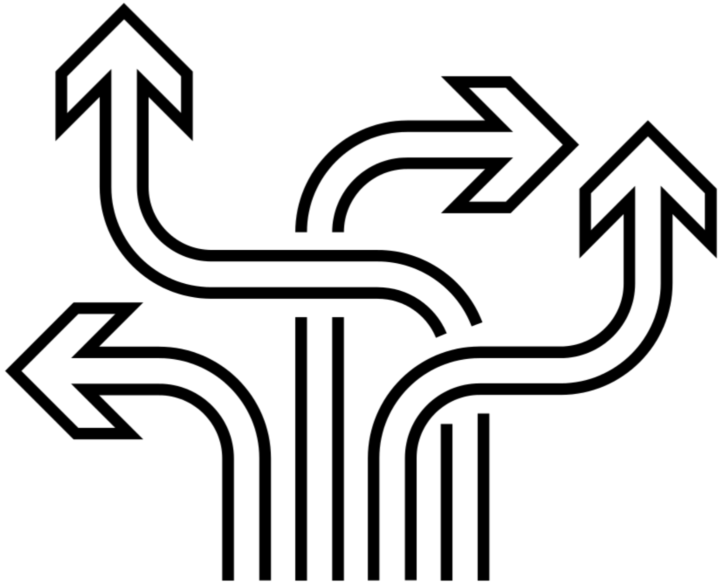
PERSONALISED HEALTHCARE – HOW CAN ML HELP?



ML FOR HEALTHCARE PERSONALISATION

How to structure the talk?

- Explain the most popular techniques
    - One click away
- Categorize
    - Type of data
        - Supervised-unsupervised techniques
    - Task
        - Diagnosis, prognosis, classification etc.
    - Other …



ML FOR HEALTHCARE PERSONALISATION

How to structure the talk?

Let the problem guide us.

Disclaimer: The choice of works presented in this tutorial is by no means an indication of preference or superiority of the method.

ML FOR HEALTHCARE PERSONALISATION

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

“ Need to *understand the patient condition*, its dynamics and provide optimal patient treatment.”

ML FOR HEALTHCARE PERSONALISATION

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

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.

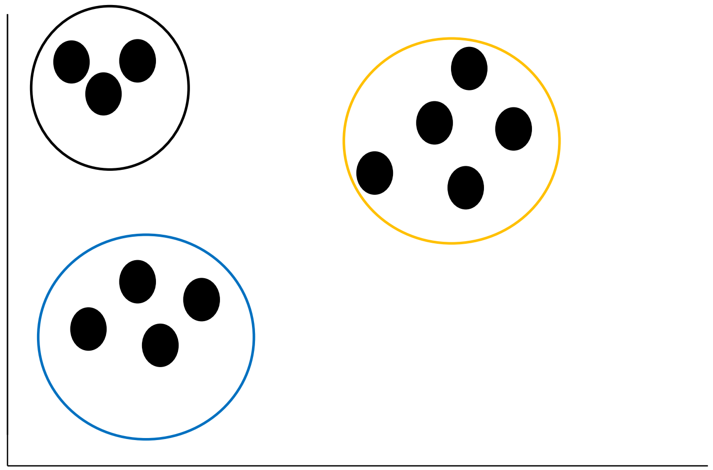
Decision Trees	K-means	Neural Networks	
Random Forests	Nearest Neighbour	Support Vector Machines	
Ensemble Methods	Hierarchical Clustering	Regression	. . .

Choice:

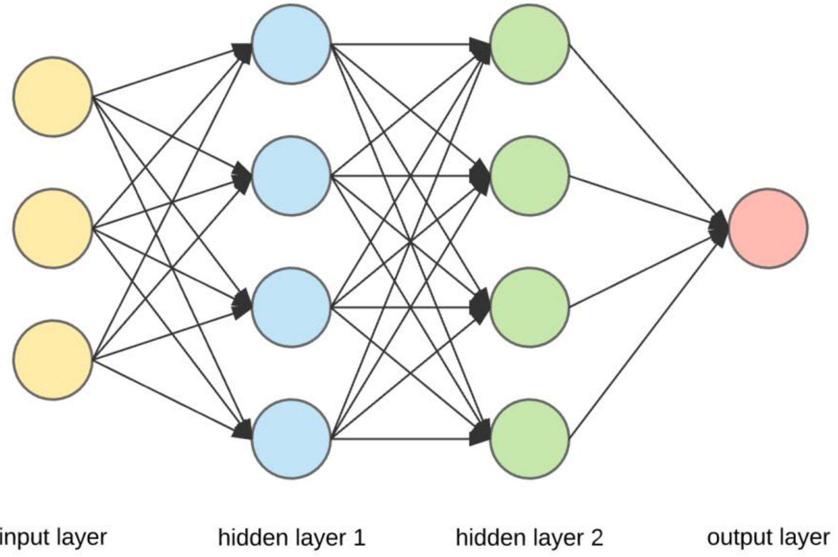
- ❑ As a first step towards understanding
- ❑ Familiarity of the user with the algorithm
- ❑ Availability of the corresponding software implementation

MODEL-FREE APPROACHES - EXAMPLE

Clustering



Neural Network



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

Patient
Vector of
symptoms

Has the
disease or
not

MODEL-FREE APPROACH – 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

Work by [**Doshi et al., 2014**]

MODEL-FREE APPROACH – 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.

Work by [**Doshi et al., 2014**]

MODEL-FREE APPROACH – 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 [_,_,…,_,_|_,_,…,_,_|. . . _,_,…,_,_| _,_,…,_,_]

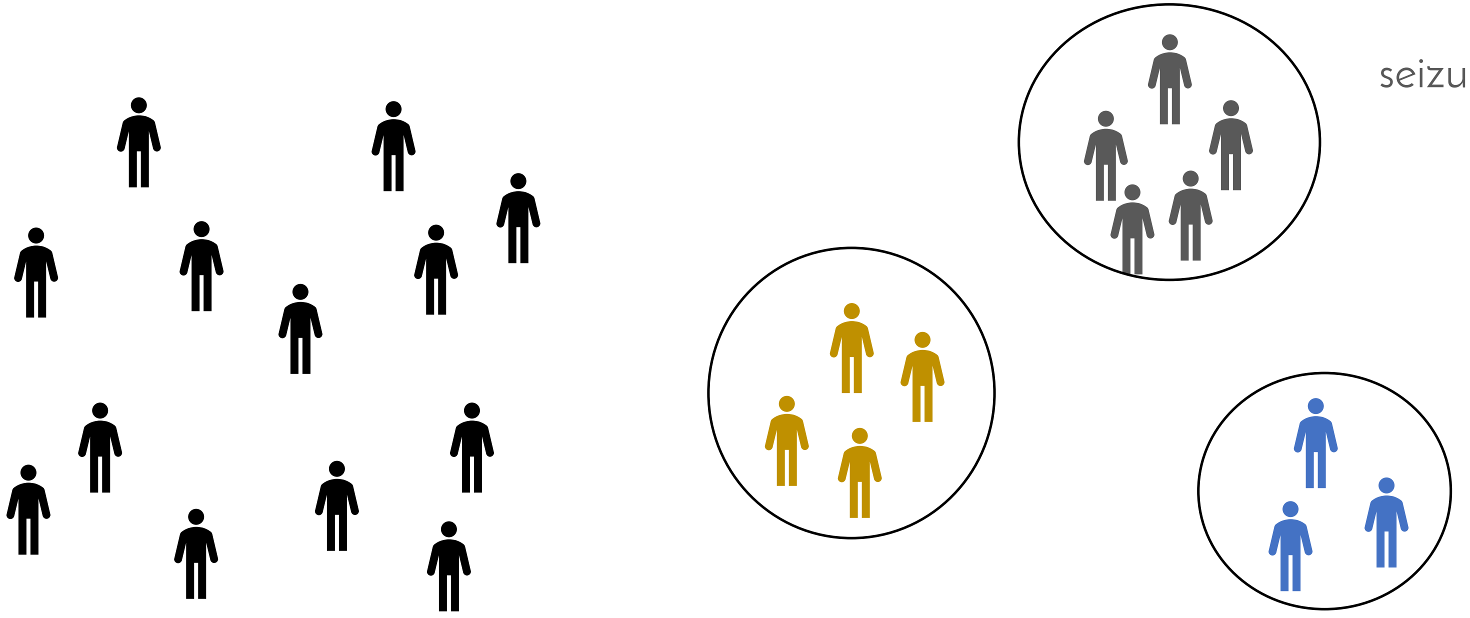
45 comorbidities

Work by [**Doshi et al., 2014**]

MODEL-FREE APPROACH – APPLICATION ON ASD

Results:

Three distinct subgroups were identified



seizures

Multisystem disorders

Psychiatric disorders

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**]

Work by [**Doshi et al., 2014**]

ML FOR HEALTHCARE PERSONALISATION

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

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

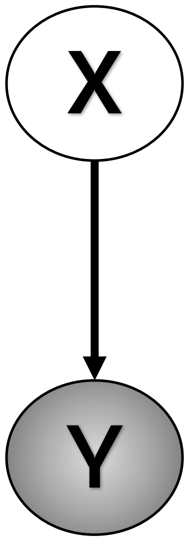
A set of assumptions (defined by the user) to describe how the observed data is generated.

Assumptions

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

Observed data
(clinical findings)

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)

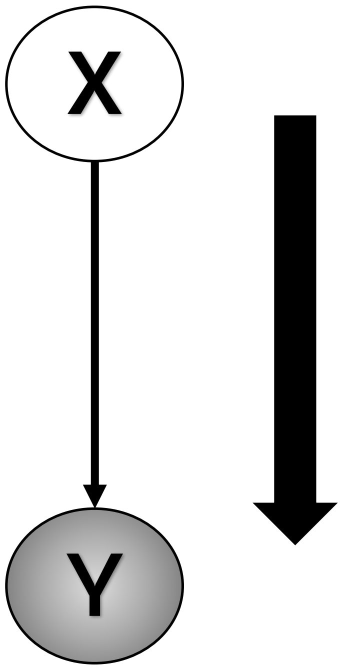
Probability theory to express all forms of uncertainty

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

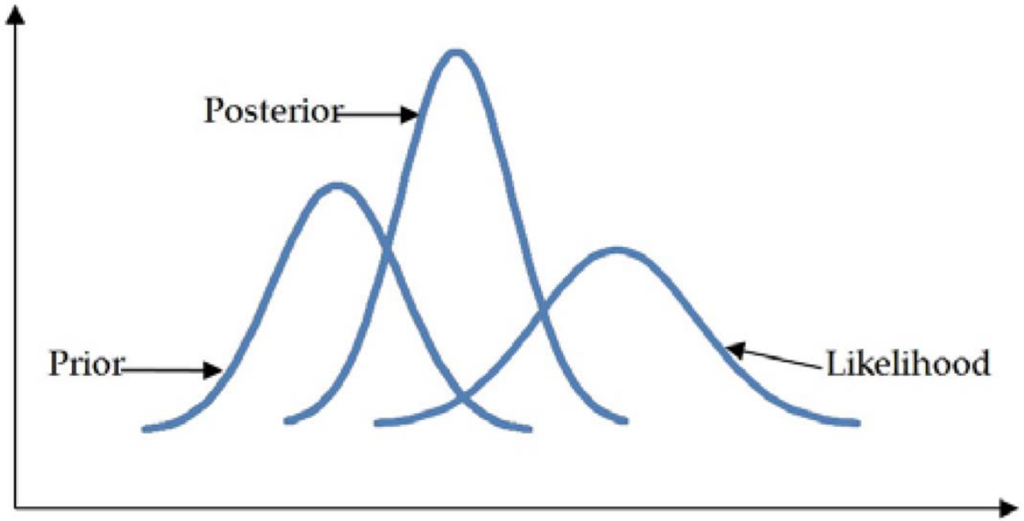
Learning: infer the value of the unknown quantities.

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.

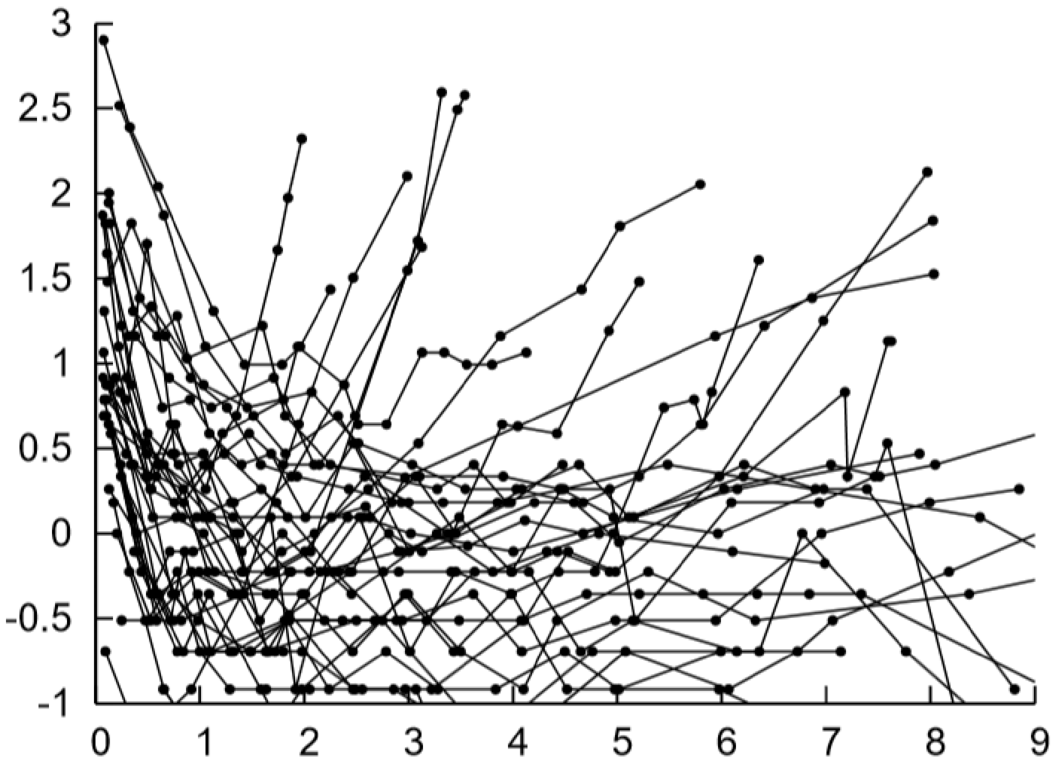
- Target:

Predict future disease trajectory

Trajectory of lung severity over time

- Challenge:

Individualize prediction by capturing variability

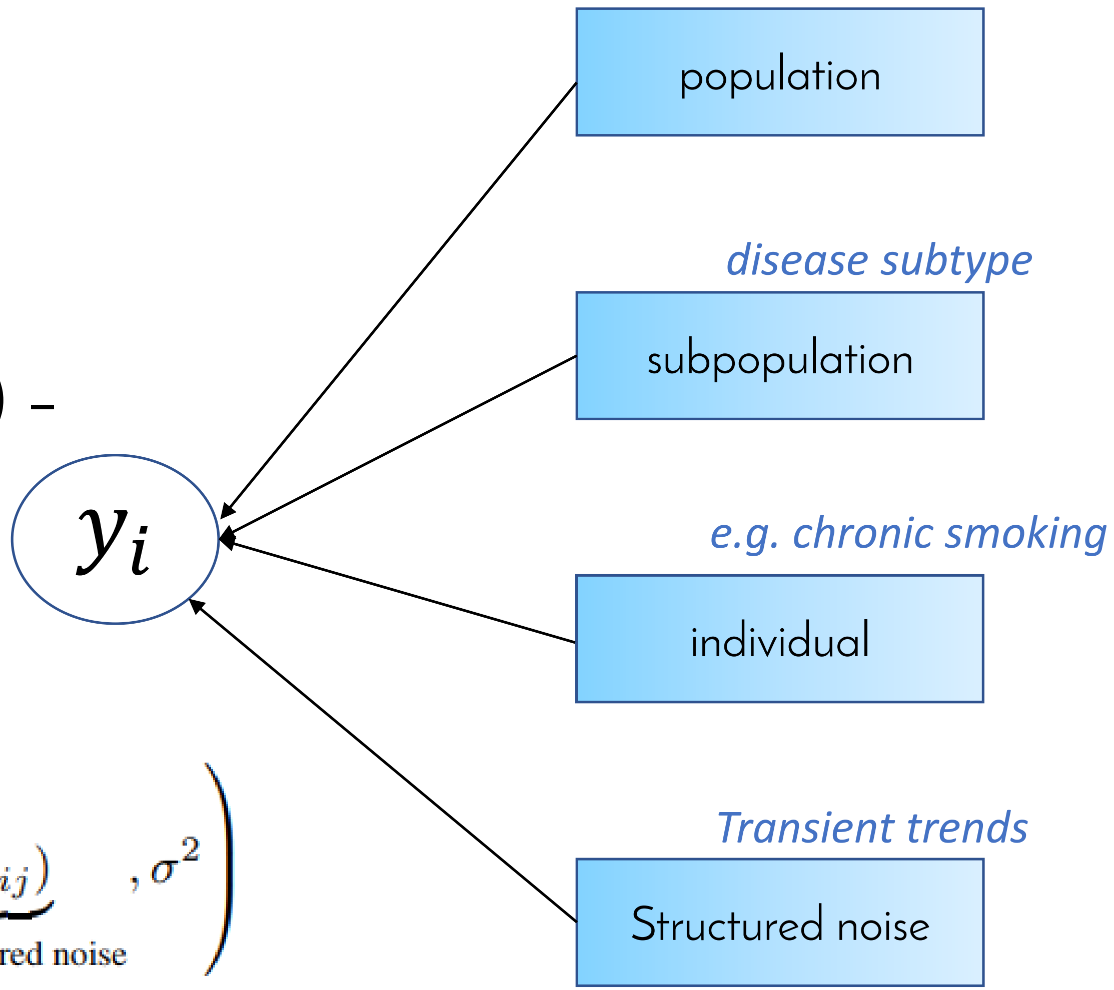


Work by [**Schulam et al.**, 2015]

Model based approach – Individualised disease progression model

- Assumptions: 4 factors of variability
- Model:

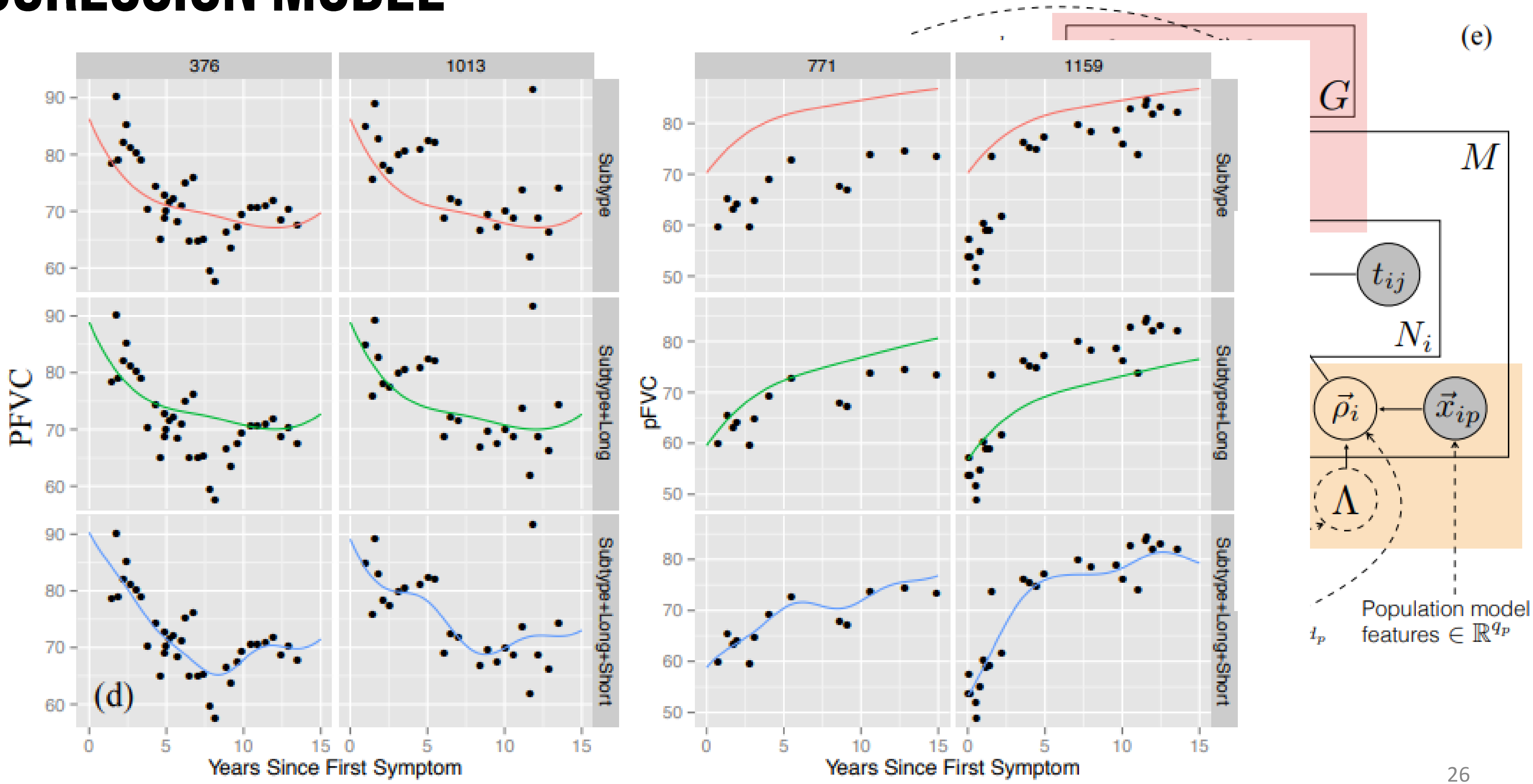
Multi-level model (Latent variable model) – organise variability in different levels



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

[**Schulam et al.**, 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

ML STRATEGIES FOR HEALTHCARE PERSONALISATION

ML for personalised treatment

ML FOR PERSONALISED TREATMENT

What treatment should I give to patient?

Ideally, we want to be confident answering this.

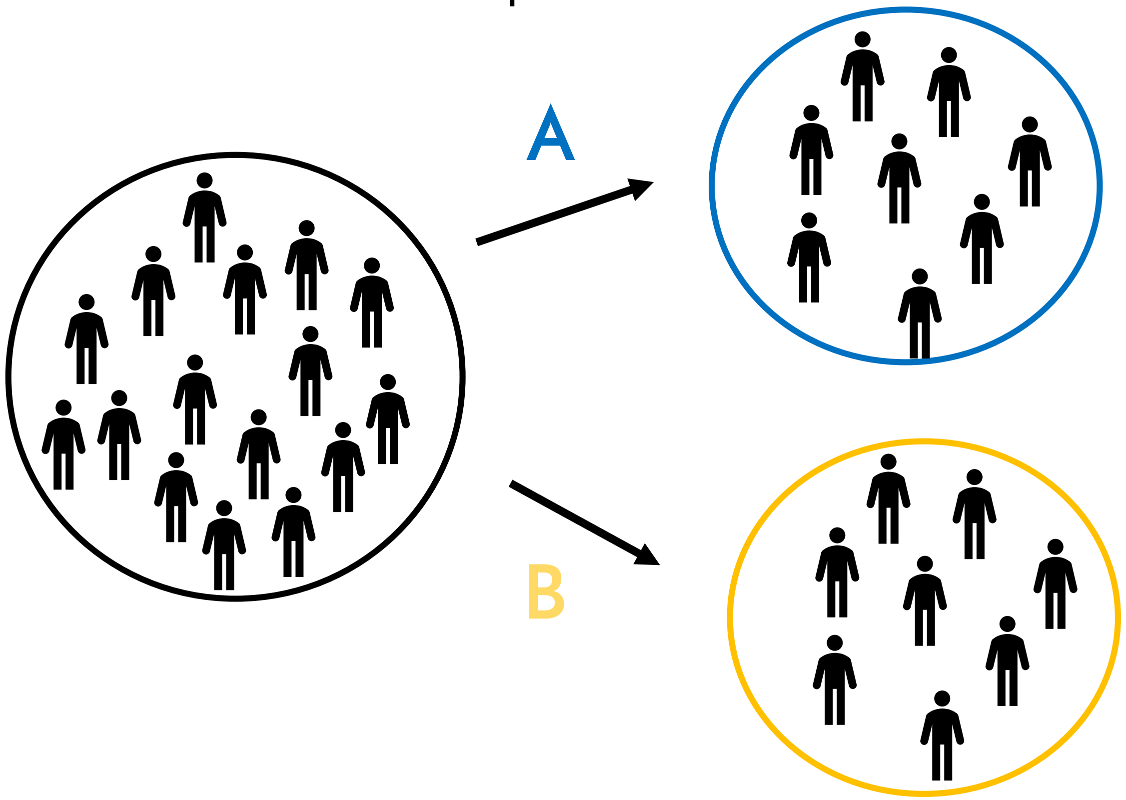
Drug
A
B

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

ML FOR PERSONALISED TREATMENT

Randomized Control Trials "Gold standard"

Control & Manipulation



Evaluate average treatment effect

BUT:

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

- Not personalised – only population effect

ML for Personalised Treatment

[Absent controlled experiment, Observational data are used]

Cheaper, Faster, in Plethora

Limitations

- Human population high heterogenous
- Doesn't contain all possible outcomes for all treatments for a patient
- Data might be biased – unknown underlying data collection protocol

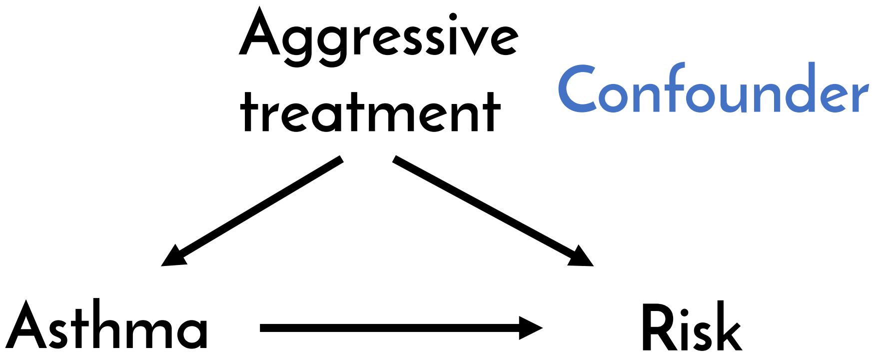
How can ML be applied on Observational data
to facilitate personalised treatment?

ML FOR PERSONALISED TREATMENT

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!



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

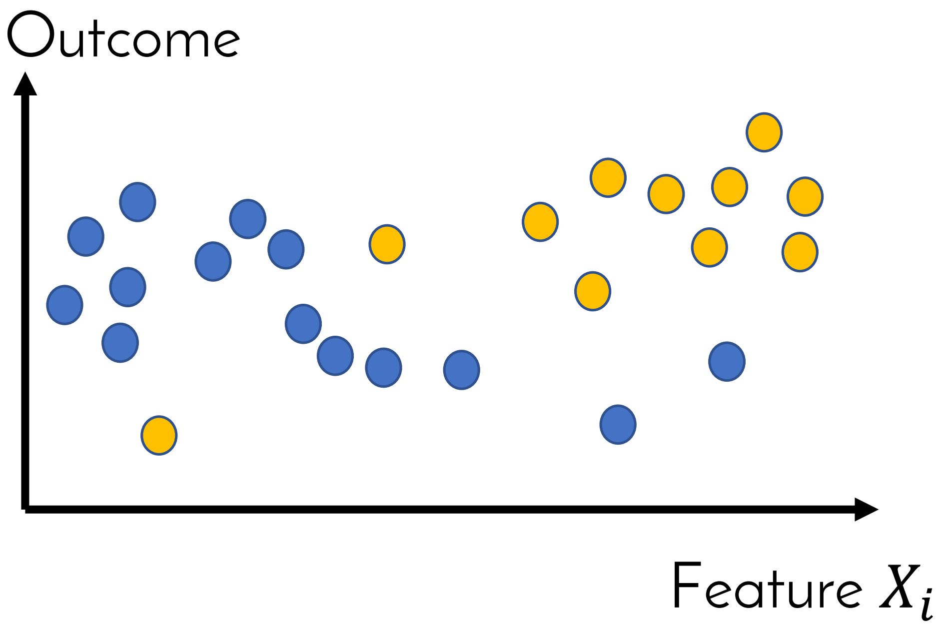
FACTUAL

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



Factuals,
A, B

Idea: Compute distribution over counterfactuals.

How: Multi-task learning problem

$X_i \longrightarrow$ **Multi-task model** $\dashrightarrow Y_i^{(A)}$, $\dashrightarrow Y_i^{(B)}$

$$Y_i^{(A)}, Y_i^{(B)} \sim GP(0, \boldsymbol{K})$$

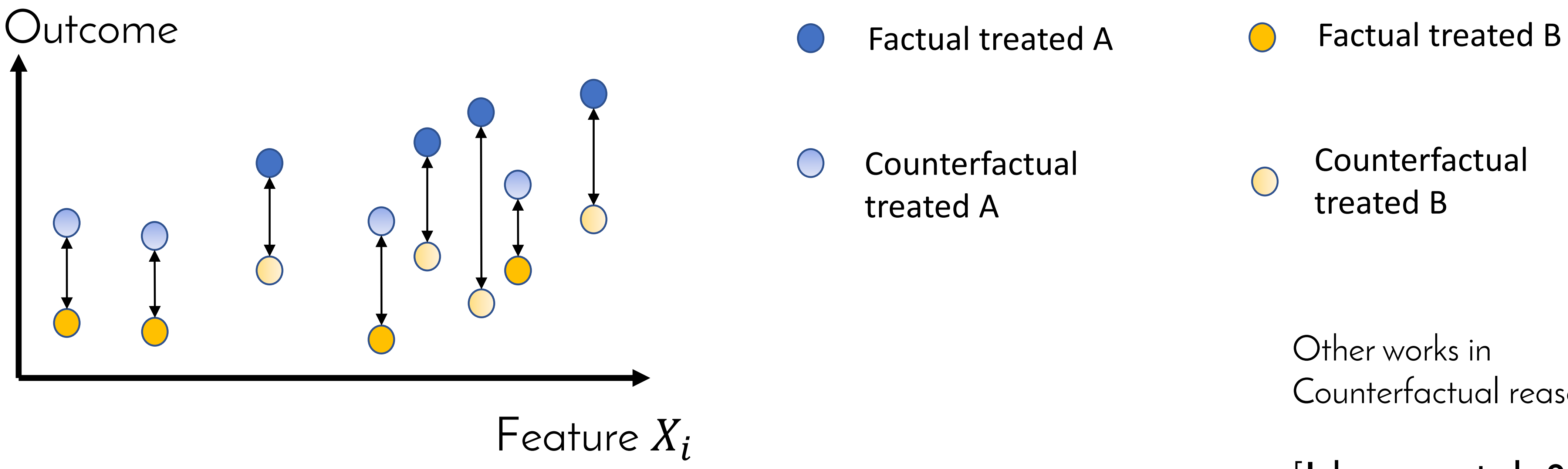
Multi-task Gaussian Process

[**Alaa et al**, 2017]

ML FOR PERSONALISED TREATMENT - COUNTERFACTUALS

[**Alaa et al**, 2017]

The Bayesian framework provides estimates of the Individualised Treatment Effect through the posterior counterfactual distribution



Other works in Counterfactual reasoning:

[**Johansson et al., 2016**]

ML Strategies for Healthcare Personalisation

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 repe**

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

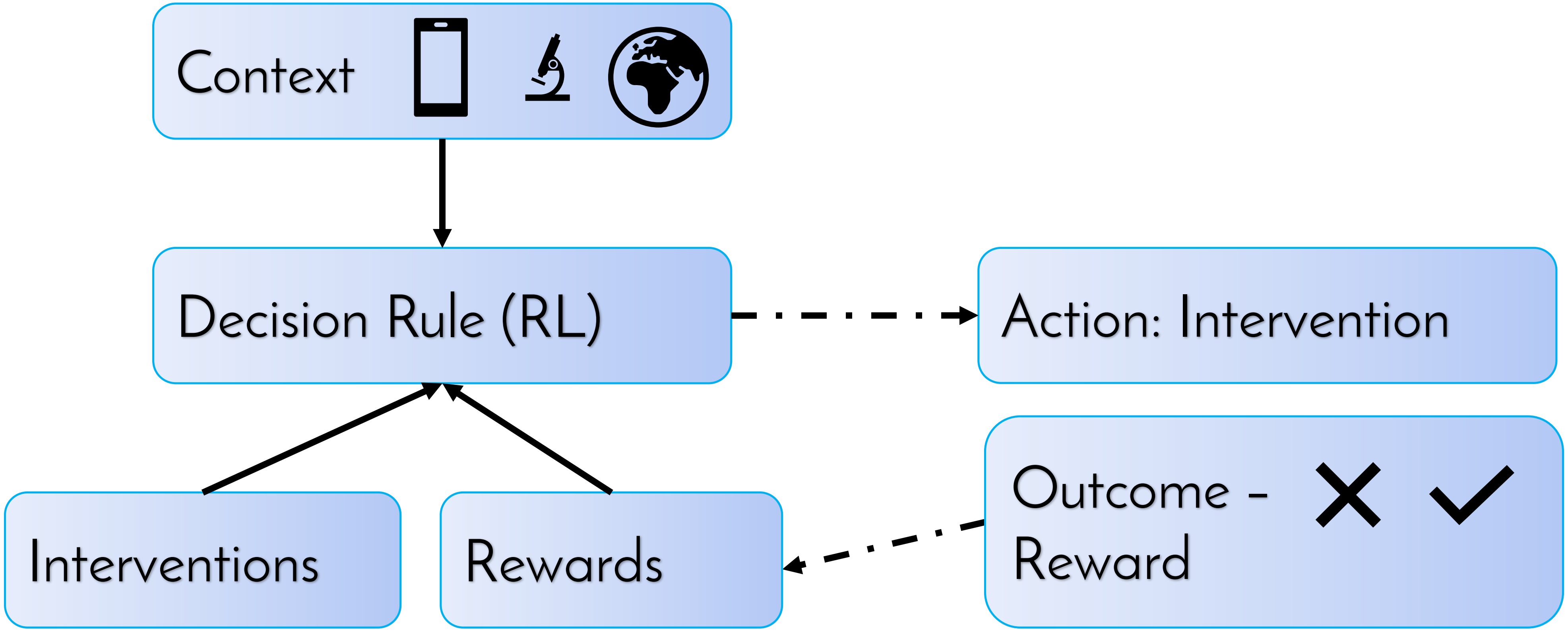
a decision rule (or policy) maps the context into an intervention option (the action)

Reinforcement learning framework + contextual bandits

Exploration - Exploitation

Personalised action

MOBILE HEALTH AND PERSONALISED INTERVENTIONS



MOBILE HEALTH AND PERSONALISED INTERVENTIONS

Encouraging physical activity of diabetes patients [**Hochberg et al., 2016**]

An intervention app to encourage physical activity

Approach: Encourage physical activity through personalised messages

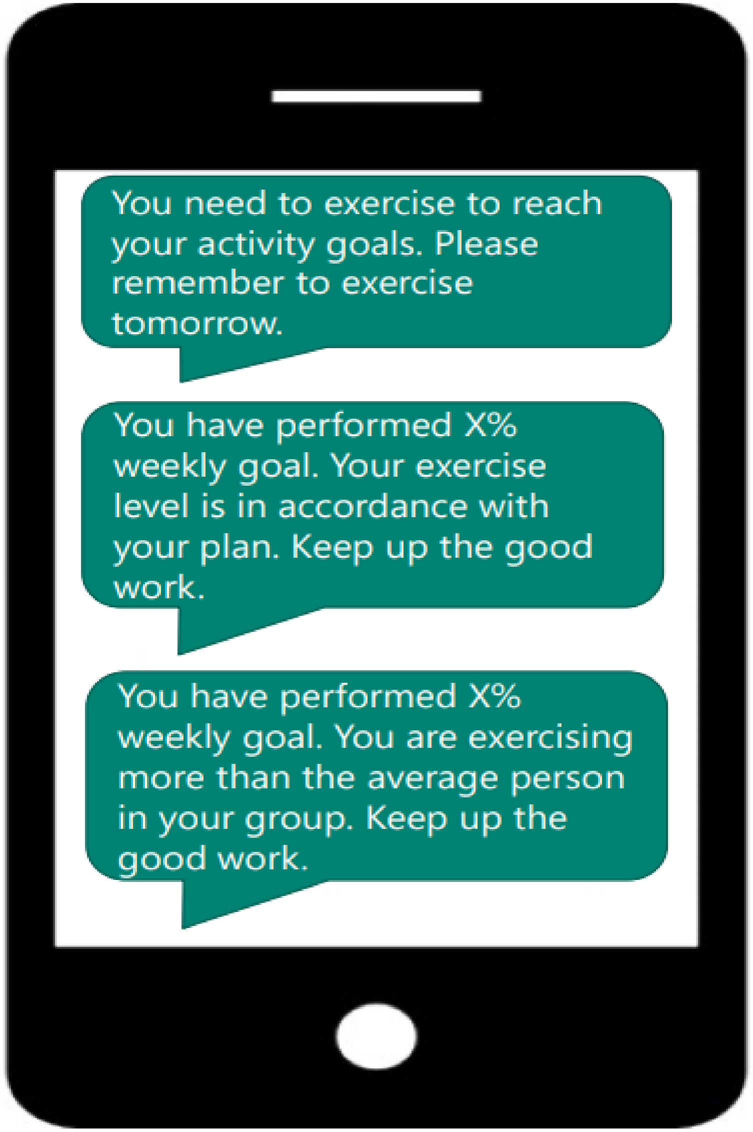
Method: RL with contextual bandits

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

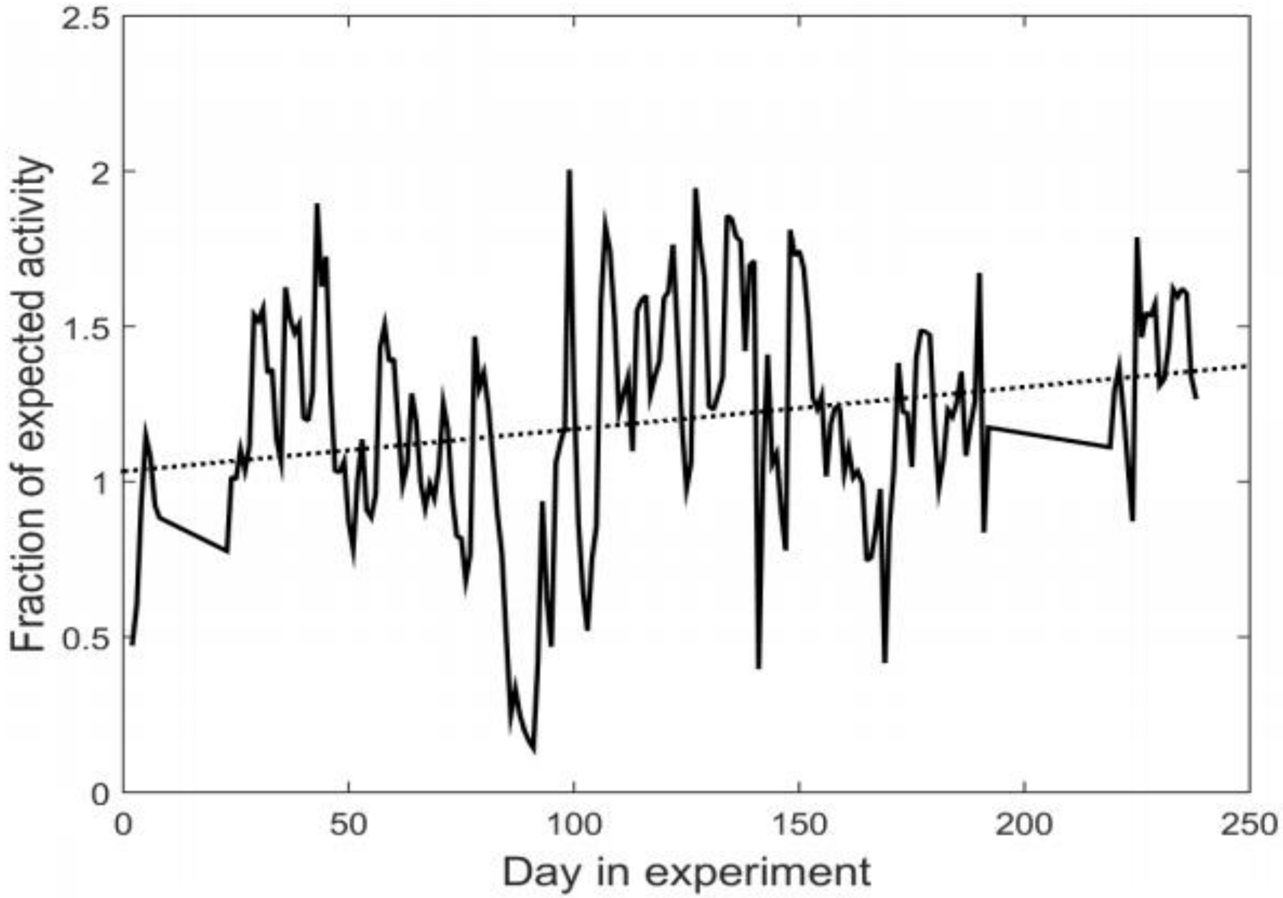
Predict respective rewards Y of [X, A]

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

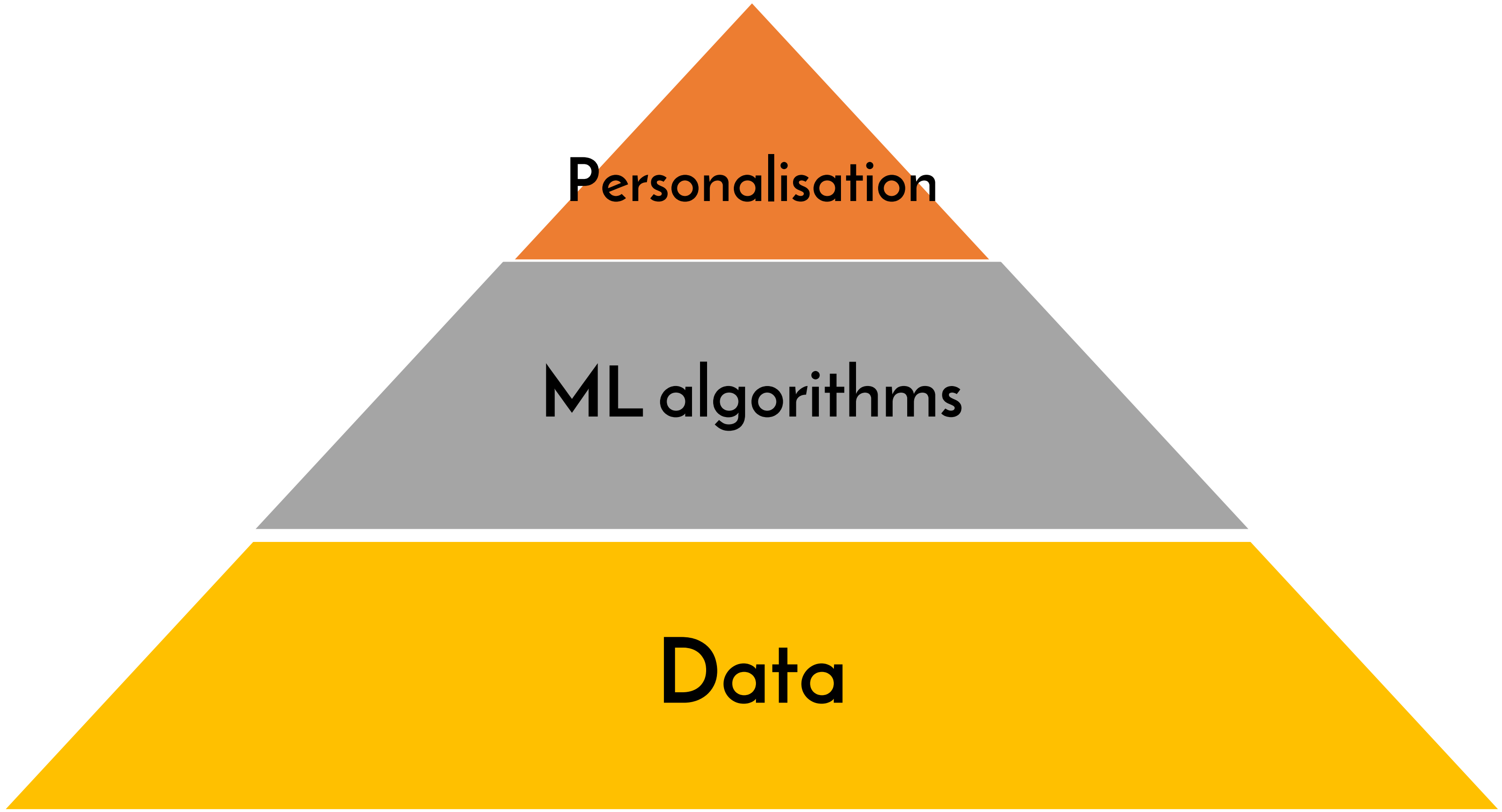
➢Psychologists, Behavioural scientists, HCI experts.

Need synergy of sciences

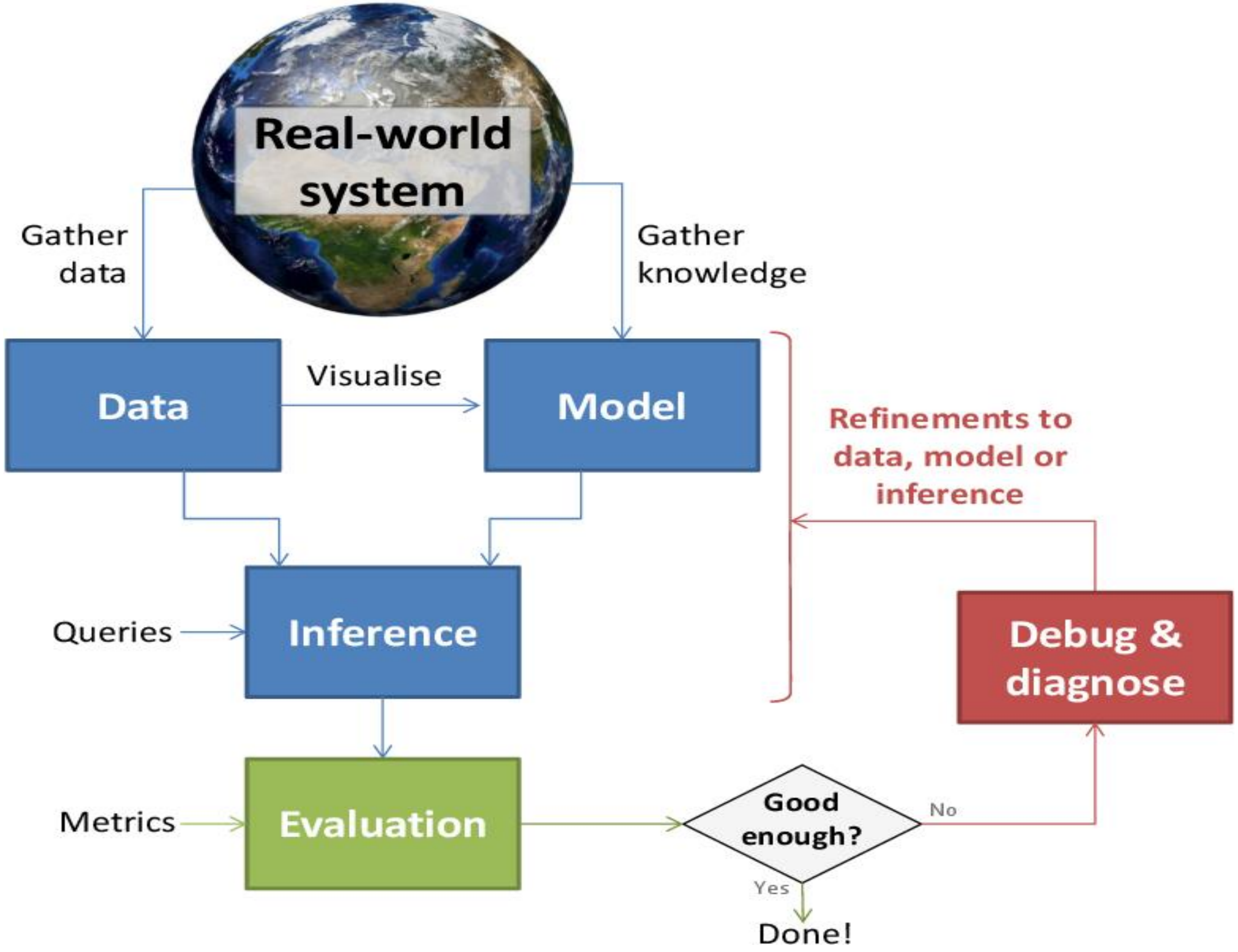
HEALTHCARE PERSONALISATION AS A THREE LEVEL PROCESS

Interconnected parts

Increased awareness at every level



LIFECYCLE OF INTELLIGENCE



[Bishop et al, 2015]

MANY THANKS TO

- Danielle Belgrave
- Zoubin Ghahramani
- Allan Tucker
- Neil Lawrence
- Sebastian Nowozin
- Aditya Nori

To all of 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 Control*. 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.

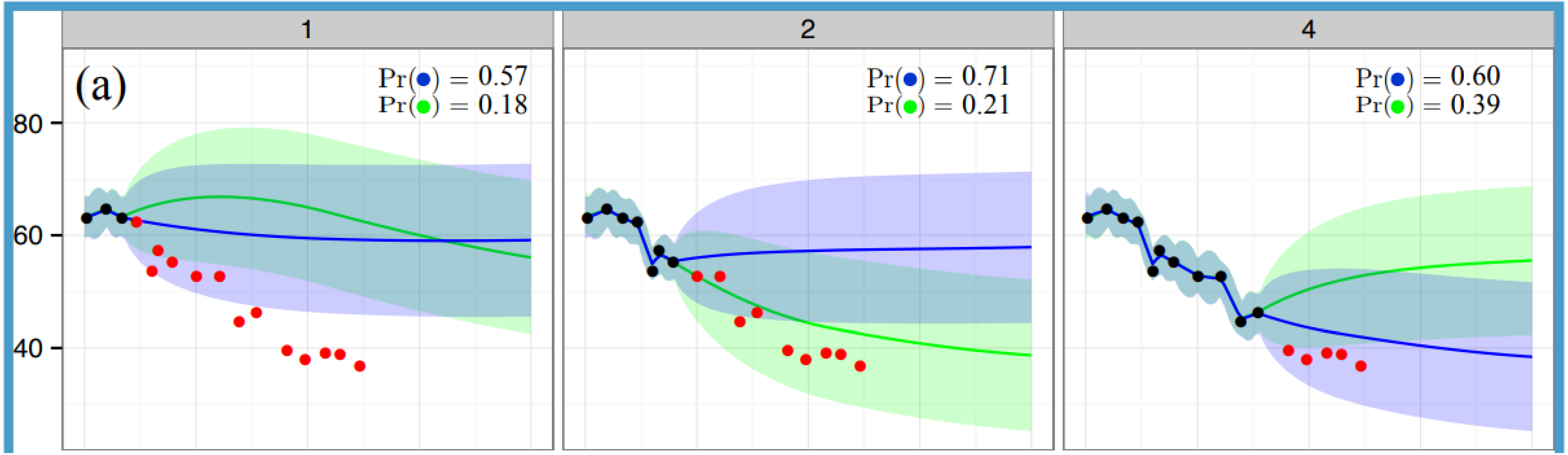
Frederik D. Johansson, Uri Shalit and David Sontag. *Learning Representations for Counterfactual Inference. ICML, 2016.*

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.

Future challenges for ML in healthcare

Lamiae Azizi

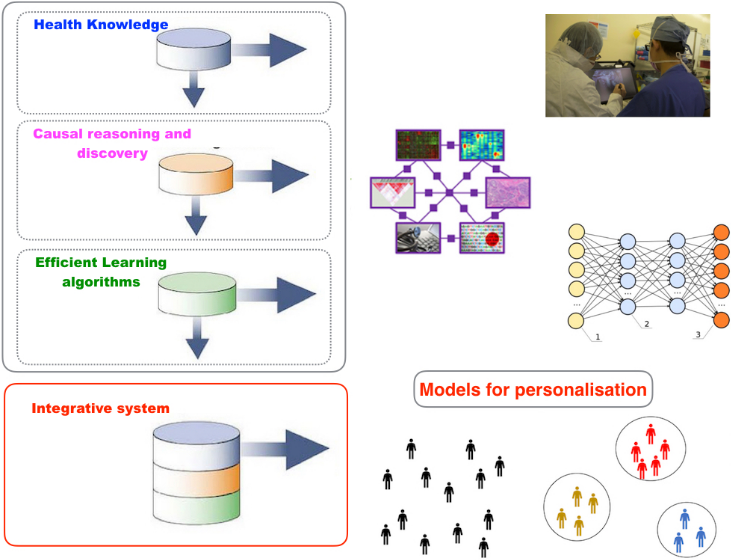
University of Sydney

10 July 2018

Plan

1. Developing the unified framework
   - Encoding the expert knowledge
   - Equipping the machinery with causal reasoning
   - Learning algorithms for complex structures
2. Rigourous Framework for trusting the model for deployment ?
3. From research to clinical implementation

Unified framework : Pillars

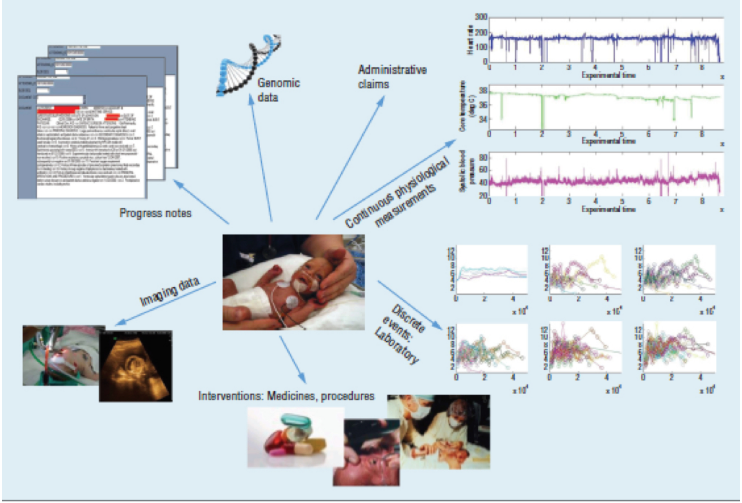


3


Health Knowledge

Health knowledge

Saria, 2014



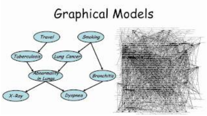
Health knowledge

Data Challenges	Technical Challenges
• Integrating multi-sources high dimensional data	• Approaches to integrate heterogeneous data
• Unstructured observational data sources	• Flexible and rich way of modelling
• Missingness in data sources	• Approaches to incorporate Mechanisms

Current approaches are not enough

Health knowledge : Integration

- Graphical model : Natural to encode domain specific relationships

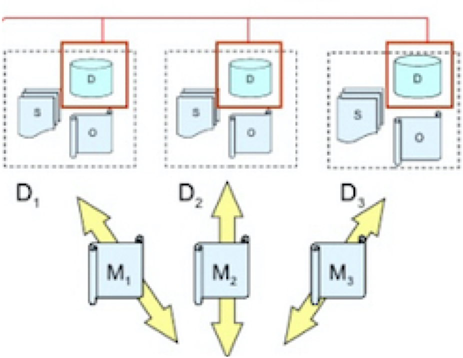


But for personalisation

Can we "even" integrate the various sources of knowledge ? !

➡ Lack of attention can lead to erroneous behavior

Health knowledge : Integration



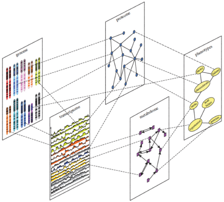
- Sources are more trusted than others
- Source Misspecification ➡ contaminate estimation and update

Principled criteria

Modular vs Full approaches combining various sources

- Integrating ”omics” and clinical
  1. Combining ”omics” still in its infancy
  2. EHR is uniquely positioned to aid when coupled with ”omics” data

No platform for EHR standardisation and ”omics” translation



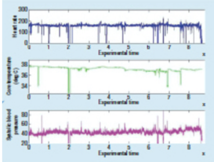
holistic system view of patient

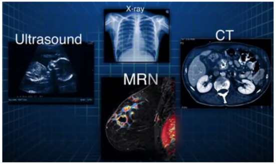
Combining genotype-phenotype, social and environmental

9

Health knowledge : "Messiness" challenge

- Continuous temporal measurements, images or text





Novel richer and flexible approaches

- Accurate for longitudinal data : inhomogeneous time series
- New memory models
  - Not evenly spaced
  - Cover long durations
  - Early events affect patient many years later

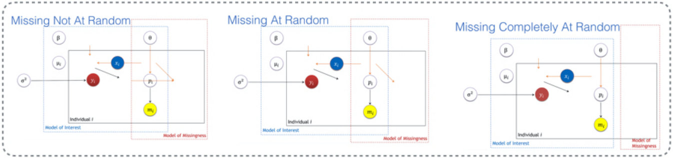
Health knowledge : "Missingness" challenge

- **Received little attention in ML**
- Sources of Missingness need to be understood
- Modelling the Missingness mechanisms

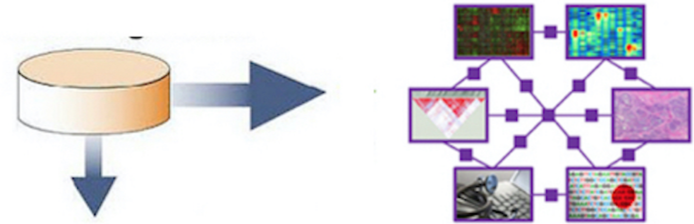
Ignoring Missingness ➜ lead to incorrect results

Unified framework

Approaches accommodating various mechanisms for various sources



Causal reasoning and discovery



Causality reasoning

- Most ML techniques lack cause-effect reasoning
- Next-generation health data : observational

Challenges for personalisation

Reasoning about learning from data through the lens of ”causal models”

- Strong assumptions
- Encoding assumptions in a compact and usable form

➡ Not a trivial matter !

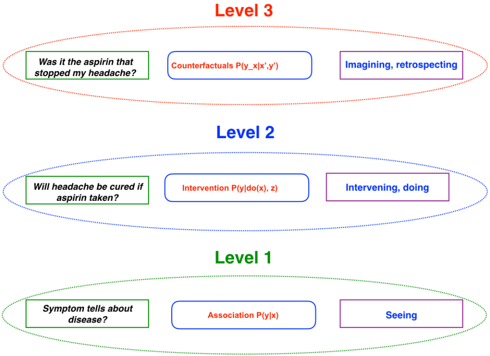
Causal inference vs causal learning

Causal discovery

- Unsupervised learning of causal relationships
- Estimate the causal structure under assumptions

➡ Challenging but promising

Counterfactual reasoning



Counterfactual reasoning, Pearl 2018

- ”Learning Machines can not answer questions about interventions non encountered”
- ”Most do not provide a representation from which answers can be derived”

Counterfactual reasoning

- If system optimises property of the observed data :
  - Back to association level ➡ No answer to ”what if”
- Complex objective functions are not an answer

Schulam et al, 2017

- Situation : Drug given to sicker patients
- Outcome : Patients die
- Model : Predicts drug kills patient (even beneficial)

Bias in the treatment policy is not considered

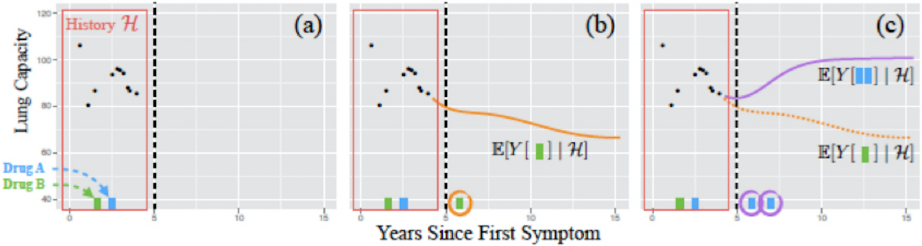
Approaches from observational data that can make

Counterfactual predictions of outcomes if an experiment run

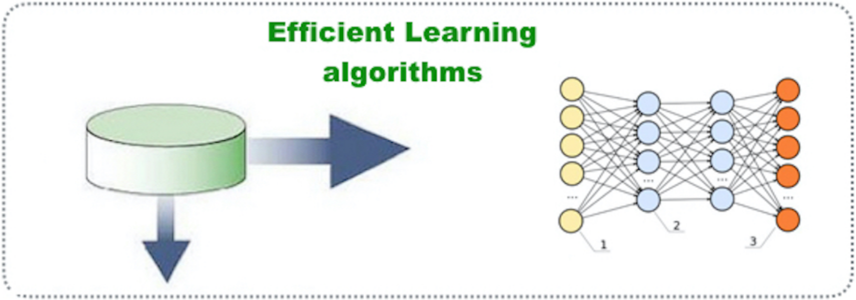
Counterfactual reasoning

Schulam et al, 2017

- Potential outcomes framework : outcomes under different actions
- Equate to counterfactual models under hypothetical interventions



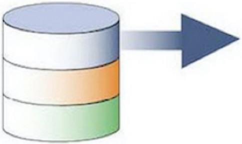
→ Promising early results in ML for healthcare


Efficient Learning
algorithms
1
2
3

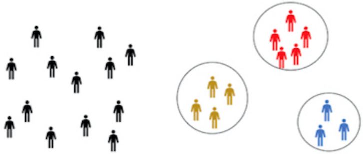
- Efficient Learning algorithms :
  - Robust approximation
  - Scalable algorithms
  - Adaptive continually learning

➡ Less challenging for ML community !

Integrative system



Models for personalisation



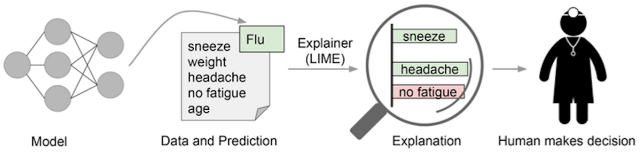
Plan

1. Developing the unified framework

2. Rigourous Framework for trusting the model for deployment ?
   - Interpretability
   - Fairness
   - Transparency, Testability and Validation

3. From research to clinical implementation

Interpretability : What does it mean ? !

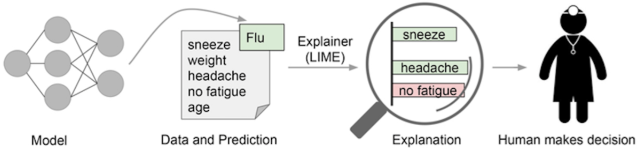


Various definitions

- Causal inference models : Interpretable models
- Feature space minimisation
- Model regularisation
- Post-hoc analysis

➡ Interpretable models : More likely to be adopted by medical practitioners

Interpretability vs Justification ? !



- Explaining a prediction vs path to the prediction explained

Ghassemi et al, 2018 and Ribeiro et al, 2016

- Identifying data points most responsible for prediction
- May help with security concerns

⚠ counter-intuitive to privacy concerns

➡ ”Justifiability” tools for the unified framework needed

Fairness : What is it and why ?

- Data quality and model choice encode unintentional discrimination
- Learning from existing clinical practice can amplify the bias

➔ Systematic disparate

Need for systems that can alert to such unwanted behaviours

Algorithmic fairness still in its infancy

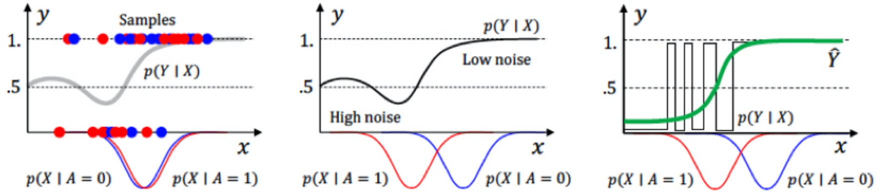
Fair model

Errors are distributed similarly across protected groups, as measured by a cost function

Fairness : What is it and why ?

Chen et al, 2018

- Fairness in prediction of an outcome Y
- Predictions are based on a :
  1. Set of covariates X : medical history of a patient in a critical care
  2. A Protected attribute A : self reported ethnicity



Which Fairness criteria and what cost

Transparency, testability and validation

- Transparency : Whether assumptions are plausible or more needed
- Testability : Whether assumptions are compatible with data
- Meaningful validation criteria : Moving beyond the current performance measures

➡ Novel criteria for validating models and assumptions

Plan

1. Developing the unified framework
2. Rigourous Framework for trusting the model for deployment ?
3. From research to clinical implementation

- Training on large dataset and assume deployment
  - ➡ Stops learning once produced
- Patient populations, recommended treatment procedures change
  - ➡ Statistical Target changes

➡ Performance degradation

Learning approaches, Ghassemi et al 2018

- Robust to changes
- Continually update

➡ Need to be considered early in systems design

Generalisability

- No guarantee for a model learned on one hospital to generalise to a new one
- Infrastructure varies across sites and health systems

ML opportunities

- Data normalisation
- Data collection at different sites

➡ Generalisability not only a modelling challenge

Clinician-Machine Interaction

Detecting individuals at risk early $\neq$ treating them early

- Systems that interact and collaborate with clinicians
- Leverage strengths of physicians and learning systems
- Having the patient and institutional preferences part of the model ? !

Increase of trust and adoption in clinical decision support

➡ Systems allowing for iterative feedback implemenation

Beyond modelling and decision support

- Augmenting Data from RCTs with observational EHD

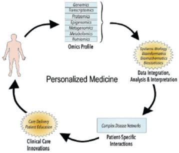
  ➡ New therapies and practice guidelines

- Novel adaptive trial designs

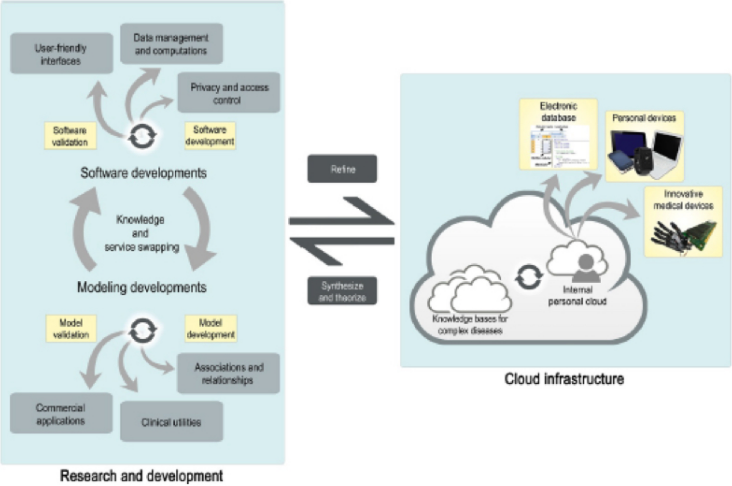
  ➡ Reducing the cost of developing new therapies

- Learn who is most likely to benefit from available ressources

  ➡ Optimizing the allocation of limited ressources



Future



User-friendly interfaces
Data management and computations
Privacy and access control
Software validation
Software development
Software developments
Knowledge and service swapping
Modeling developments
Model validation
Model development
Associations and relationships
Commercial applications
Clinical utilities
Research and development
Refine
Synthesize and theorize
Electronic database
Personal devices
Innovative medical devices
Internal personal cloud
Knowledge bases for complex diseases
Cloud infrastructure

References

- S. Saria. **A $3 Trillion Challenge to Computational Scientists: Transforming Healthcare Delivery**, August 2014. IEEE Intelligent Systems. Vol. 29, Issue 4
- P. E. Jacob, L.M. Murray, C. Holmes, C.P. Robert. **Better together? Statistical learning in models made of modules (2017)**
- J. Pearl. **Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution (2018).**
- P. Schulman, S. Saria. **Reliable Decision support using Counterfactuals Models.** NIPS 2017.
- M. Ghassemi, T. Naumann, P. Schulman, A.L. Beam, R Ranganath. **Opportunities in Machine Learning for Healthcare. (2018)**
- M.T. Ribeiro, S. Singh, C. Guestrin. **"Why Should I Trust You?": Explaining the Predictions of Any Classifier.** International conference on Knowledge Discovery and Data mining (KDD). ACM, 1135-1144.
- I Chen, FD Johansson, D Sontag. **Why Is My Classifier Discriminatory? (2018)**

Thank you !