Machine Learning Applications for Health

Dr Brian Chapman
brian.chapman@unimelb.edu.au
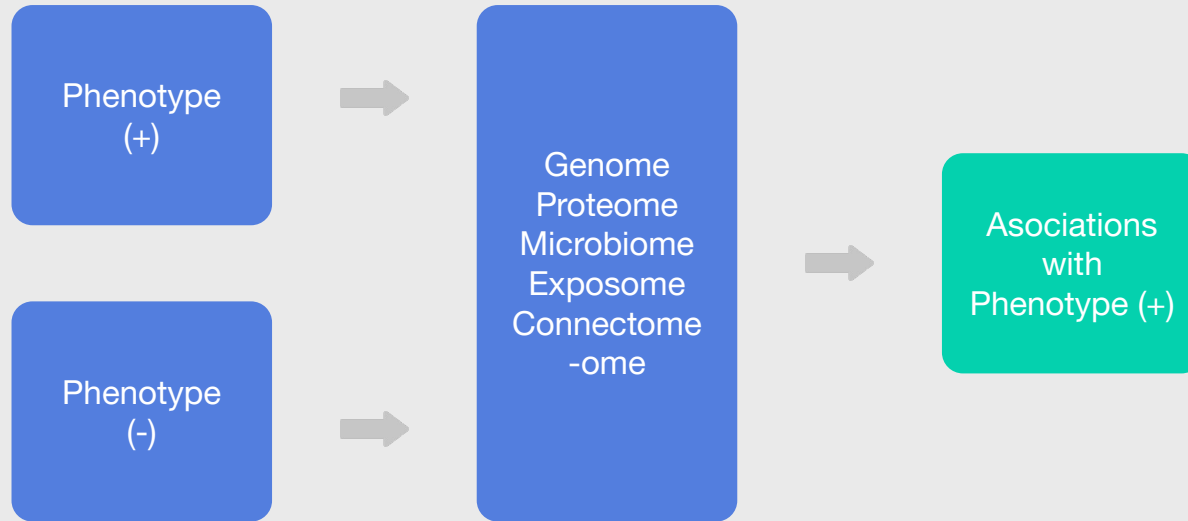
Dr Daniel Capurro
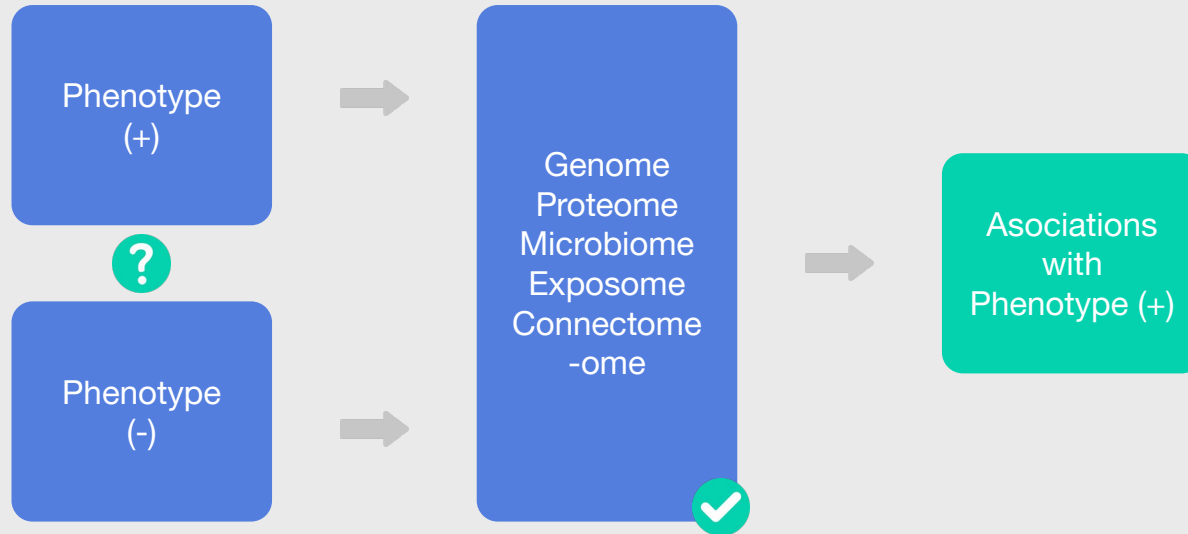dcapurro@unimelb.edu.au

Digital Phenotypes

Phenotype

- Combination of observable attributes of an organism

- Result from the combination of genotype + environment

Typical Association Study

Typical Association Study

Extended Phenotype

- Concept introduced by Dawkins in 1983

- Phenotype should also include the effect we exert on the environment

- Digital environments and automated data collection make possible the digital phenotype
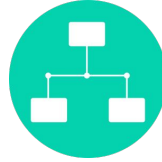
Digital Phenotypes

- Machine Learning
- Retrospective EHR studies

- Quality measurement

- Patient recruitment

Phenotyping Methods

- Manual chart abstraction

- Diagnostic Codes (ICD-10, CPT)

- Natural Language Processing

- Ad-hoc algorithms

Ad-hoc algorithms: some issues

- Multiple iterations over time

- Limited reuse

- Different levels of abstraction coexist in the same database

- Relative temporal queries

symptoms (1)

codes (special codes) (2)

prescription (3)

medication

Temporal Abstraction-Based Digital Phenotyping

CLINICALTIME

Objectives

- Describe clinical intervals and their use in retrospective studies

- Create a generic method to build digital phenotypes based on clinical temporal intervals.

Methods

- Systematically reviewed ICU observational studies using EHR data

- Obtained generalizable knowledge regarding interval types and their relations

- Using MIMIC Database, we built and validated phenotypes
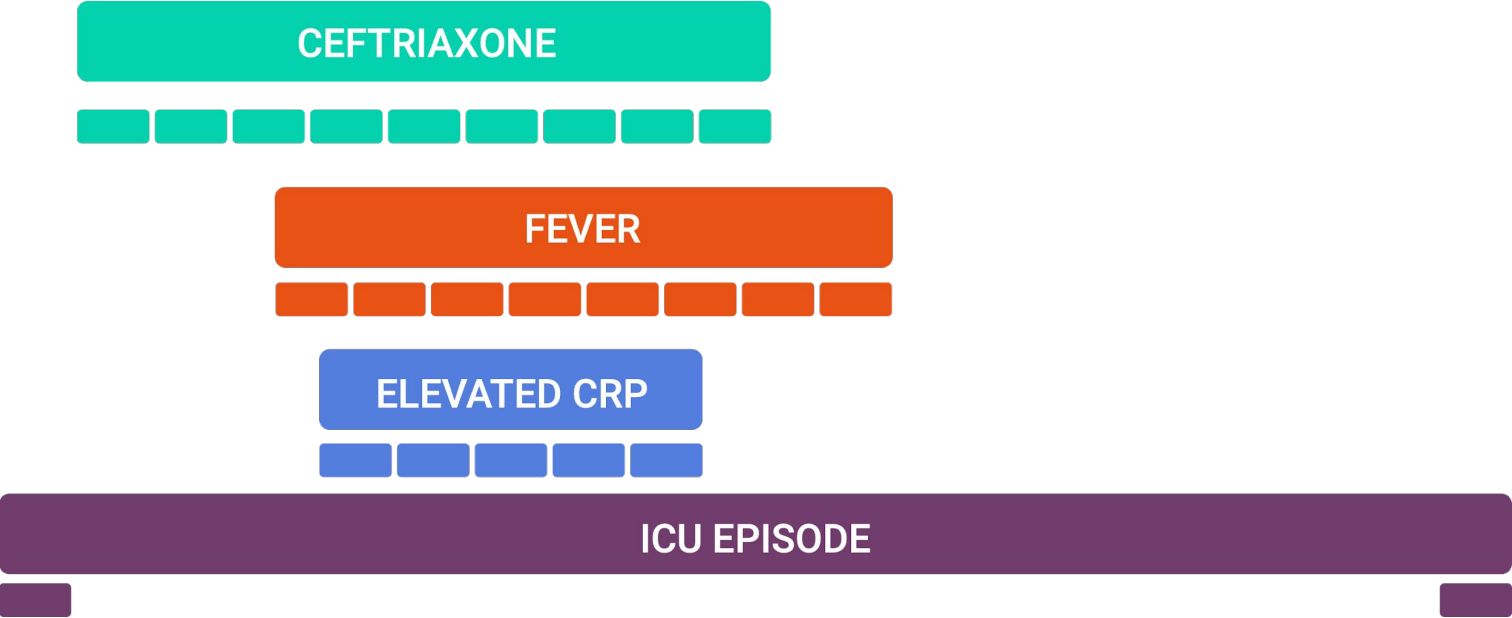
Selection and enrollment

Patients meeting all inclusion criteria and no exclusion criteria (detailed in Table 1) are candidates for inclusion in the study. In brief, patients with a high suspicion of VAP (by clinical and radiological signs), have been on mechanical ventilation at least for 96 hours or less than 96 hours if they have previously received antibiotic treatment for at least five days and have an in-hospital stay of more than seven days. The setting of the study will be the ICUs of public hospitals in Spain, Italy and Greece.

Antibiotics

Mechanical Ventilation

Hospital Stay

Rosso-Fernández C, Garnacho-Montero J, Antonelli M, Dimopoulos G, Cisneros JM. Safety and efficacy of colistin versus meropenem in the empirical treatment of ventilator-associated pneumonia as part of a macro-project funded by the Seventh Framework Program of the European Commission studying off-patent antibiotics: study protocol for a randomized controlled trial. Trials. 2015 Dec;16(1):102.

Example of Temporally Abstracted Phenotype: Ceftriaxone-induced fever in an ICU patient

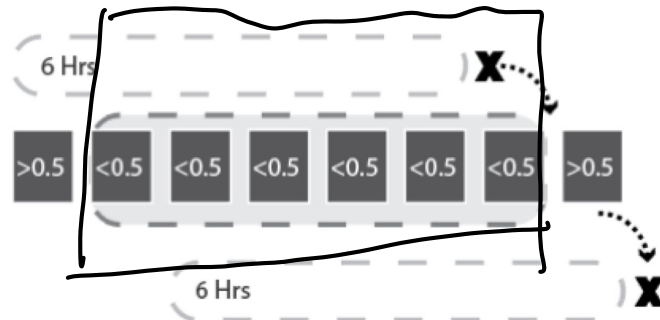# Results: types of intervals

区间类型

a. NORMAL

Patient 1: >37ºC >37ºC >37ºC <37ºC

Patient 2: >37ºC >37ºC <37ºC <37ºC

b. BOUNDED

Admission Discharge

c. MOVING WINDOW   Assignment

6 Hrs   X

>0.5 <0.5 <0.5 <0.5 <0.5 <0.5 <0.5 >0.5

6 Hrs   X

Results: temporal relations

- Thirteen qualitative temporal relations previously described by Allen

- Quantitative relations between intervals and within intervals

Relation	Symbol	Symbol for inverse
X before Y	$<$	$>$
X equal Y	$=$	$=$
X meets Y	m	mi
X overlaps Y	o	oi
X during Y	d	di
X starts Y	s	si
X finishes Y	f	fi

Allen JF. Towards a general theory of action and time. Artificial intelligence. 1984 Jul 1;23(2):123-54.

Laboratory evidence of infection
Gram stain, culture

Infection Suspected

New Antimicrobial agent started and continued for at least 4 days

AND

WBC < 4,000 or > 12,000 or > 10% bands

OR

Temperature <36ºC or >38ºC

Increase in daily minimum FiO$_2$ by 0.2

or

Increase in daily minimum PEEP values in >= 3 cmH$_2$O

Baseline Period of Stability
for at least 48 hours

Mechanical Ventilation
for at least 48 hours

VAP

IVAC

VAC
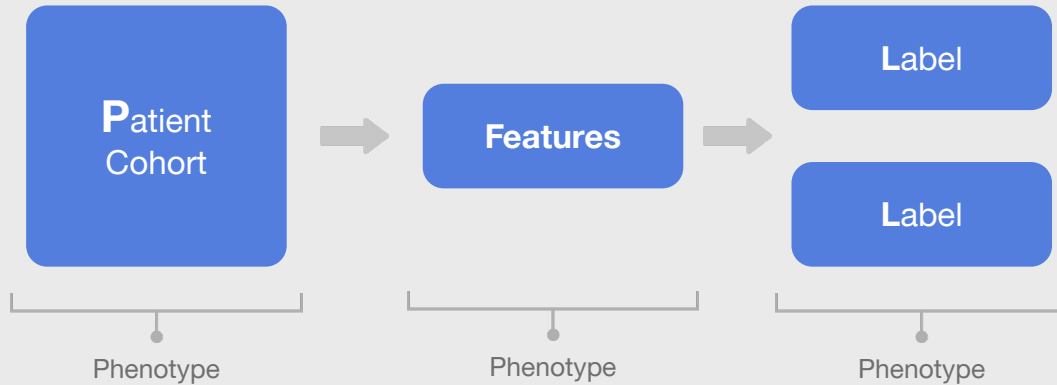
Phenotype
Building
Example

CLINICAL**TIME**

Phenotypes everywhere!!
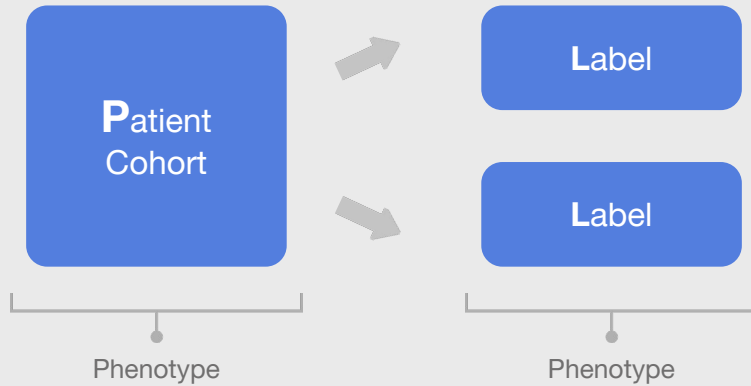
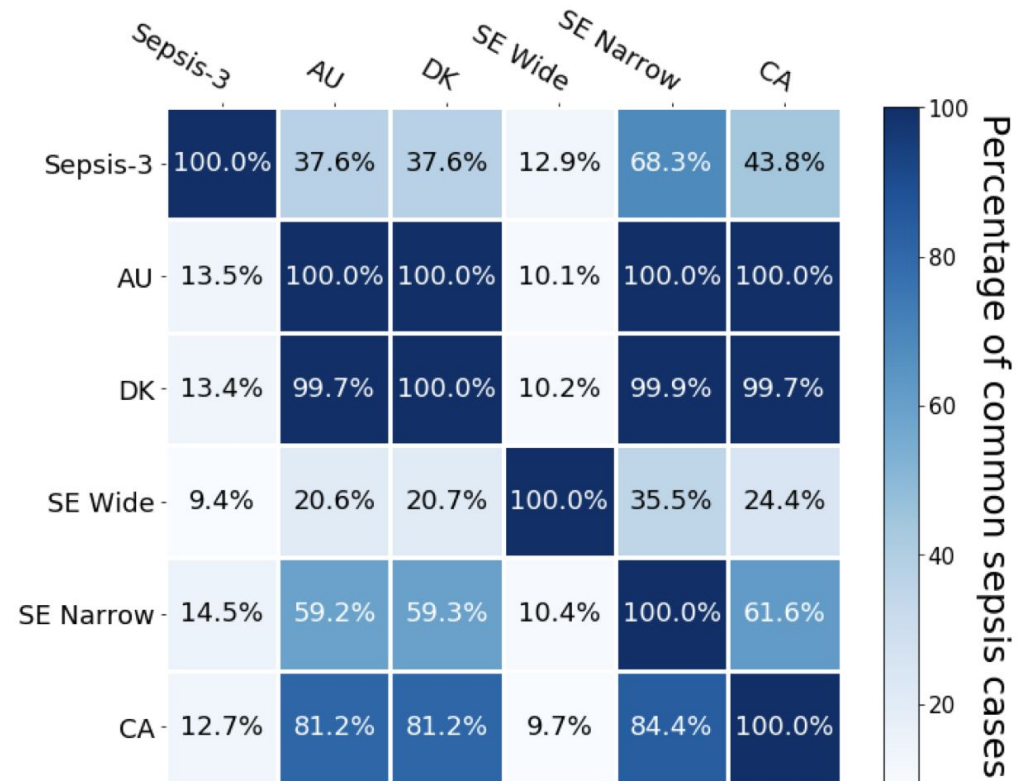# Phenotypes everywhere!!
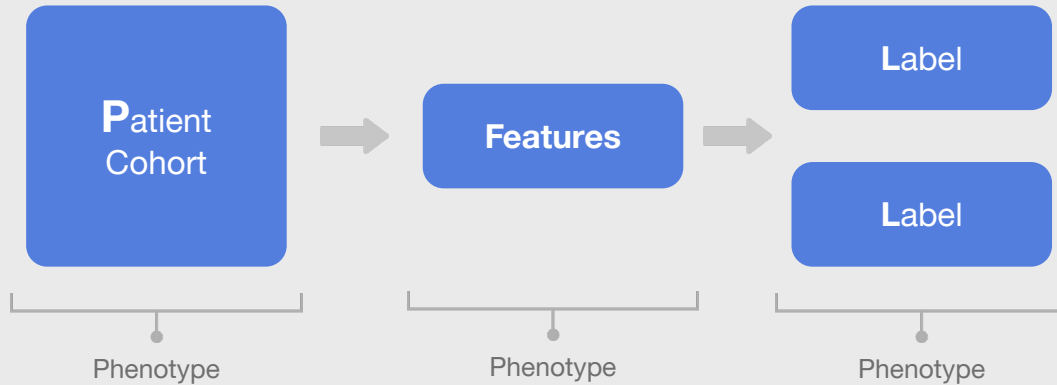
Phenotypes everywhere!!

Fig. 1. Confusion matrix of administrative coding approaches against Sepsis-3 definition as the ground true.

Table 1. Measurement of sepsis prediction performance

	Sepsis-3		AU		DK		SE Wide		SE Narrow		CA	
	AUROC	MCC	AUROC	MCC	AUROC	MCC	AUROC	MCC	AUROC	MCC	AUROC	MCC
LR	0.80	0.60	0.82	0.64	**0.83**	**0.67**	0.76	0.54	0.82	0.63	0.83	0.66
RF	**0.82**	**0.64**	0.78	0.56	0.80	0.61	0.66	0.33	0.78	0.57	0.80	0.61
GB	**0.82**	**0.64**	0.78	0.55	0.79	0.59	0.70	0.41	0.77	0.55	0.80	0.60

LR=Logistic Regression, RF=Random Forest, GB=Gradient Boosting.

Getting the phenotypes right is MORE important than the ML model you choose!!

Sources of Phenotypes

- Biomedical Literature - Don't reinvent the wheel!!!

- [EMERGE Network](#)

- [OHDSI](#)