The CALIBER Research Platform
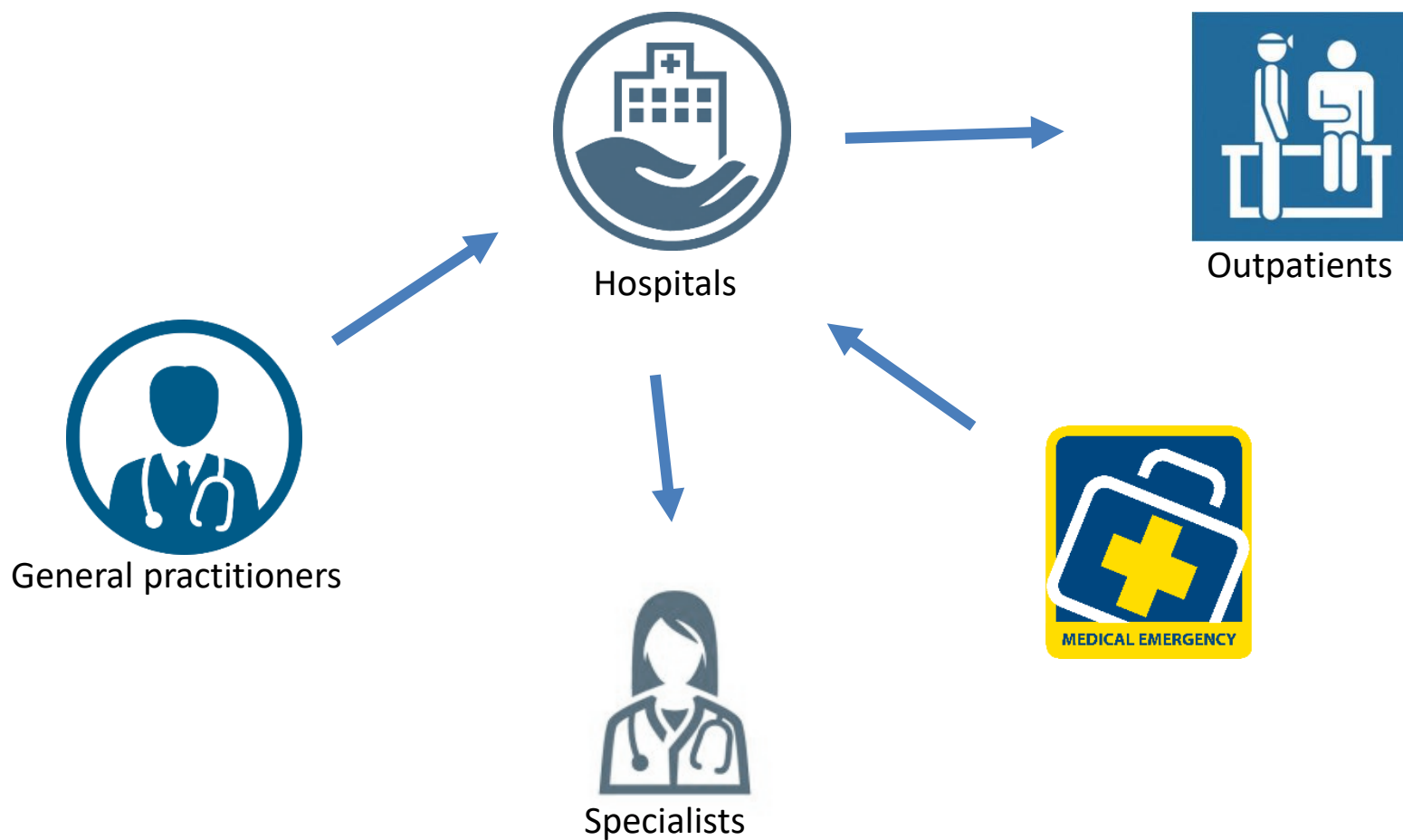## Using large-scale linked electronic health records for research

Dr Arturo González-Izquierdo
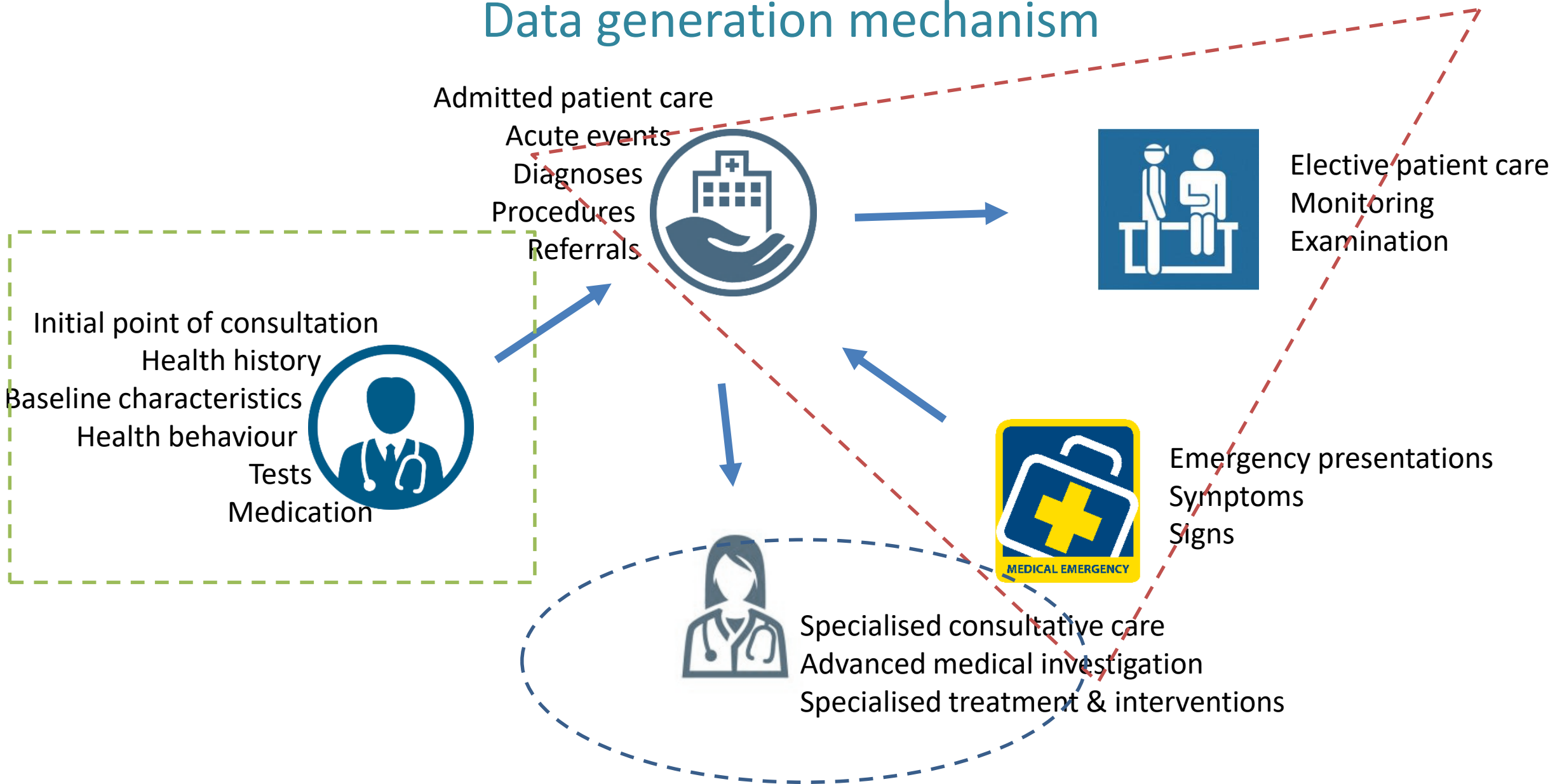
University College London

Institute of Health Informatics

- Data generation mechanism

- Linked electronic health records

- EHR phenotyping

- Challenges & opportunities

Healthcare system

Data generation mechanism

Admitted patient care
Acute events
Diagnoses
Procedures
Referrals

Elective patient care
Monitoring
Examination

Initial point of consultation
Health history
Baseline characteristics
Health behaviour
Tests
Medication

Emergency presentations
Symptoms
Signs

Specialised consultative care
Advanced medical investigation
Specialised treatment & interventions

Healthcare settings and data custodians

CPRD: GP Data

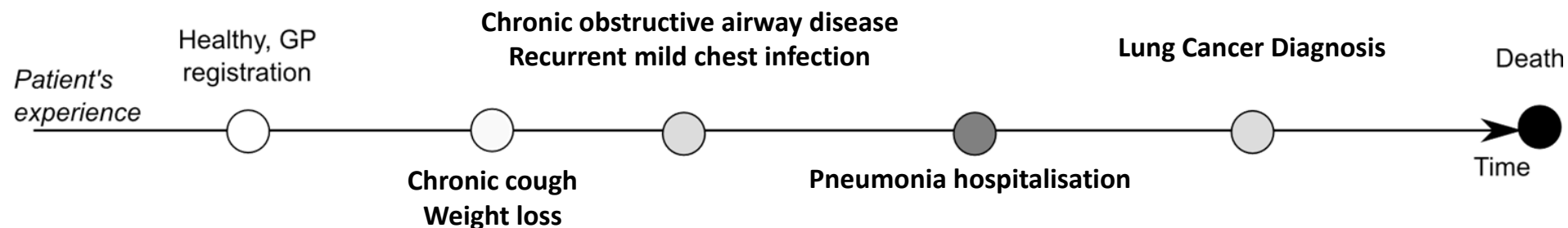**NHS Digital**: Hospital Data

Disease Registries: Tertiary care data

ONS: Mortality Data
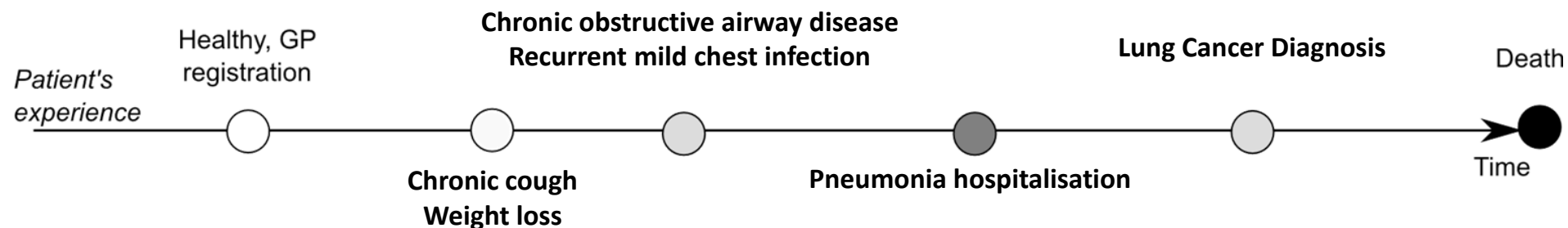
Electronic Health Records

Clinical Practice Research Datalink

Hospital Episode Statistics

National Cancer Registration and Analysis Service

Office for National Statistics

Linked Electronic Health Records

Linked Electronic Health Records

Patient's experience

Healthy, GP registration

Chronic obstructive airway disease
Recurrent mild chest infection

Chronic cough
Weight loss

Pneumonia hospitalisation

Lung Cancer Diagnosis

Death

Time

Registration date	Consultation date	Consultation date	Admission date	Admission date	Date of death
Date of birth	Blood tests (routine)	Sputum tests	Diagnosis	Diagnosis (cancer type, stage, metastases)	Underlying cause
Blood pressure	Spirometry	Diagnosis	Additional diagnosis (deep vein thrombosis Pneumonia)		Subsidiary causes
Weight	Chest x-rays	Procedures	Procedure for biopsy of lesion (bronchoscopy, Chest drain)	Procedures (surgery)	
Height	CT chest		Chest imaging (x-ray, CT, PET CT, chest drain insertion)	Treatment (chemotherapy, radiotherapy)	
Physical activity	Treatment (antibiotic, Inhalers)		Discharge date	Discharge date	
Health history (heart, diabetes, stroke)					
Smoking					
Alcohol					
Contraception					
Immunisations					

Linked Electronic Health Records

Patient's experience

Healthy, GP registration

Chronic obstructive airway disease
Recurrent mild chest infection

Chronic cough
Weight loss

Pneumonia hospitalisation

Lung Cancer Diagnosis

Death

Time

Registration date
Date of birth
Blood pressure
Weight
Height
Physical activity
Health history
(heart, diabetes, stroke) **Coded**
Smoking
Alcohol
Contraception
Immunisations

Consultation date
Blood tests (routine)
Spirometry
Chest x-rays
CT chest
Treatment (antibiotic, Inhaler) **Coded**

Consultation date
Sputum tests
Diagnosis **Coded**
Procedures

Admission date
Diagnosis
Additional diagnosis
(deep vein thrombosis **Coded**
Pneumonia)
Procedure for biopsy
of lesion (bronchoscopy, **Coded**
Chest drain)
Chest imaging (x-ray,
CT, PET CT,
chest drain insertion)
Discharge date

Admission date
Diagnosis (cancer
type, stage,
metastases)
Procedures (surgery)
Treatment (chemotherapy, **Coded**
radiotherapy)
Discharge date

Date of death **Coded**
Underlying cause
Subsidiary causes

EHR phenotype

Biometrics, test results, time dependent thresholds

```
IF HT = 1, THEN HT_comp = 1

For non-diabetic patients:
IF there are at least three occurences of (SYS_BP  >= 140 mmHg OR DIAS_BP >= 90 mmHg) within 1 year of releva
nt date) OR if there are at least three occurences of  (BP_cat = 3) within 1 year of relevant date
THEN HT_comp = 1


For diabetic patients:
IF there are at least three occurenc    f (SYS      = 130 mm OR DIAS_BP >= 80 mmHg) within one year of relevan
t date OR if there are at least thr    uren      (BP_cat = 3) within 1 year of relevant date
THEN HT_comp = 1
```
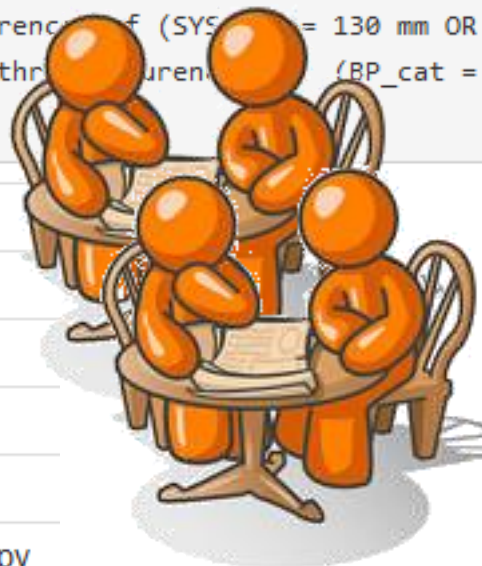
Diagnoses

6627.00	Go
6628.00	Po
662F.00	Hyp
662G.00	Hypertensive treatm.changed
662O.00	On treatment for hypertension
662b.00	Moderate hypertension control
662c.00	Hypertension six month review
662d.00	Hypertension annual review
662r.00	Trial withdrawal of antihypertensive therapy
7O01.00	High cost hypertension drugs

Medication

```
All primary care prescriptions from BNF chapters:
2.2.1 Thiazides and related diuretics
2.2.3 Potassium-sparing diuretics and aldosterone antagonists
2.2.4 Potassium-sparing diuretics with other diuretics
2.4 Beta-adrenoceptor blocking drugs
2.5 Hypertension and heart failure
2.6.2 Calcium-channel blockers
```

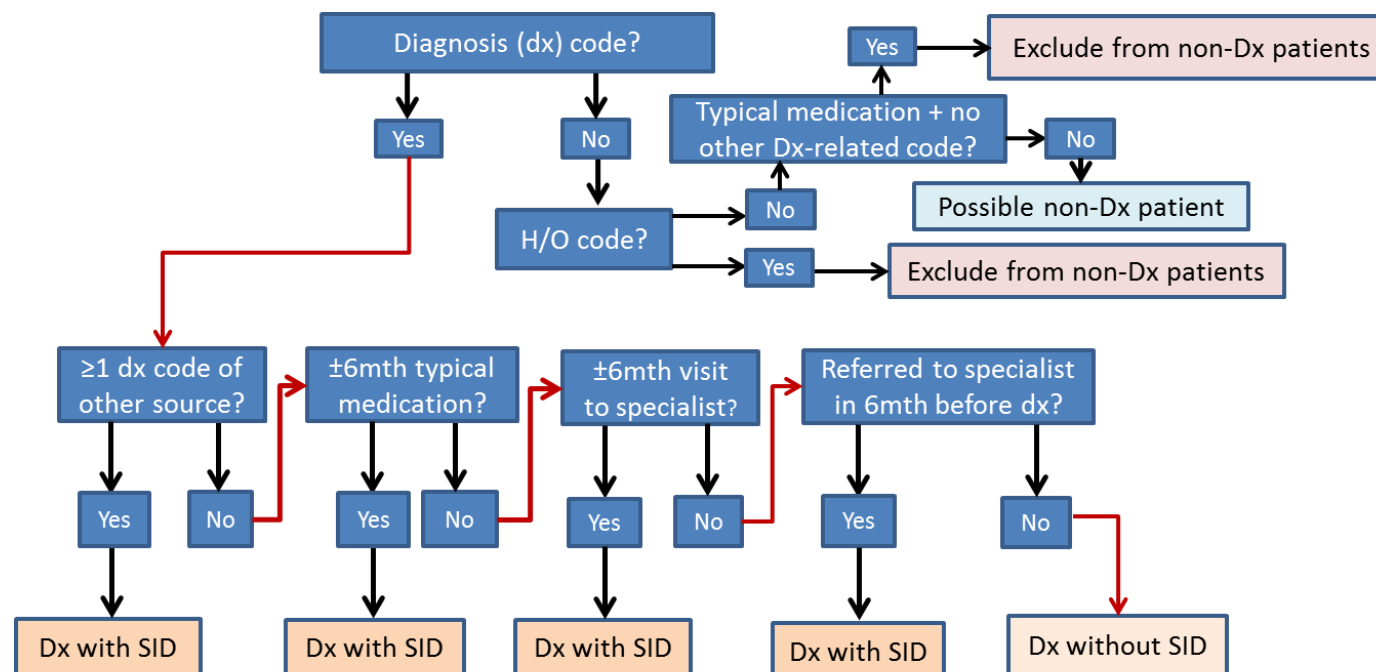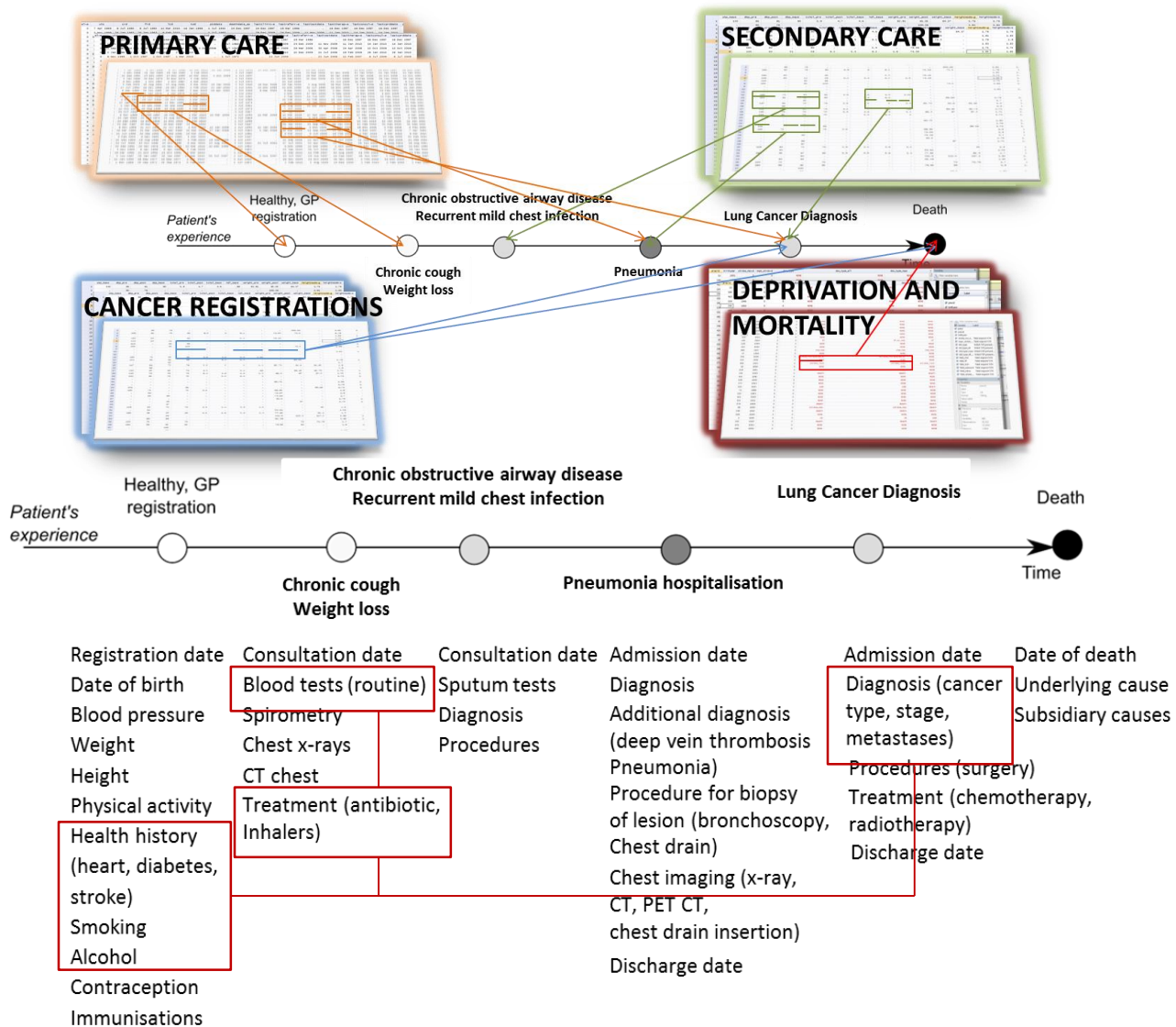## Health care utilisation patterns

EHR phenotype

Phenotype (classification)			
Binary	...		Detailed
Yes	"Definite"	Severe	
		Moderate	↑
		Mild	⋮
	Probable	Probable	↓
	Possible	Possible	
No	No	No	

EHR phenotype

- ## Extraction – Algorithm (generic)

Pujades-Rodriguez M. (2016) Heart, 102:383-398

The CALIBER Research Platform

Cohort identification methods

Deep phenotyping algorithms

Patient Population
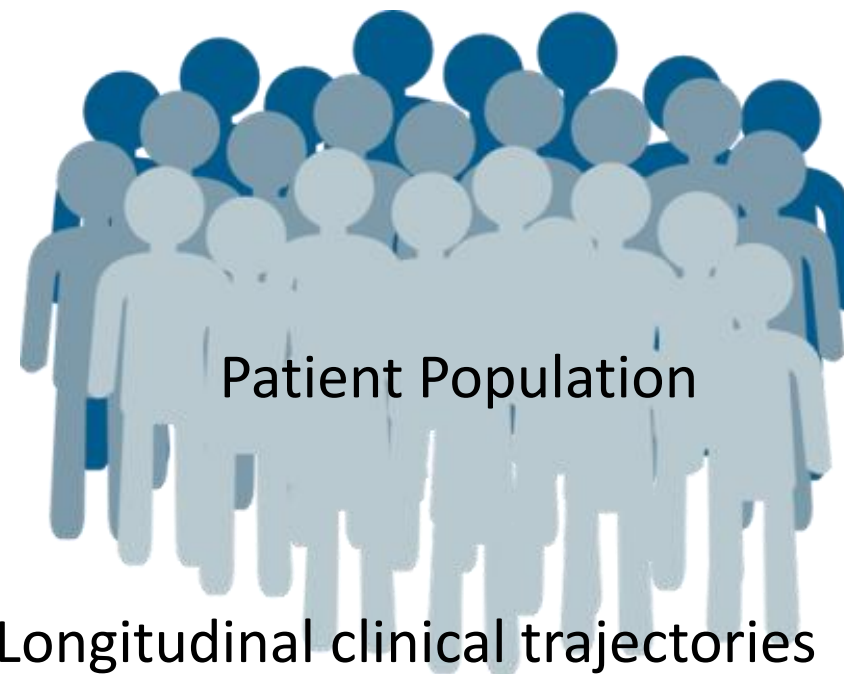
Longitudinal clinical trajectories

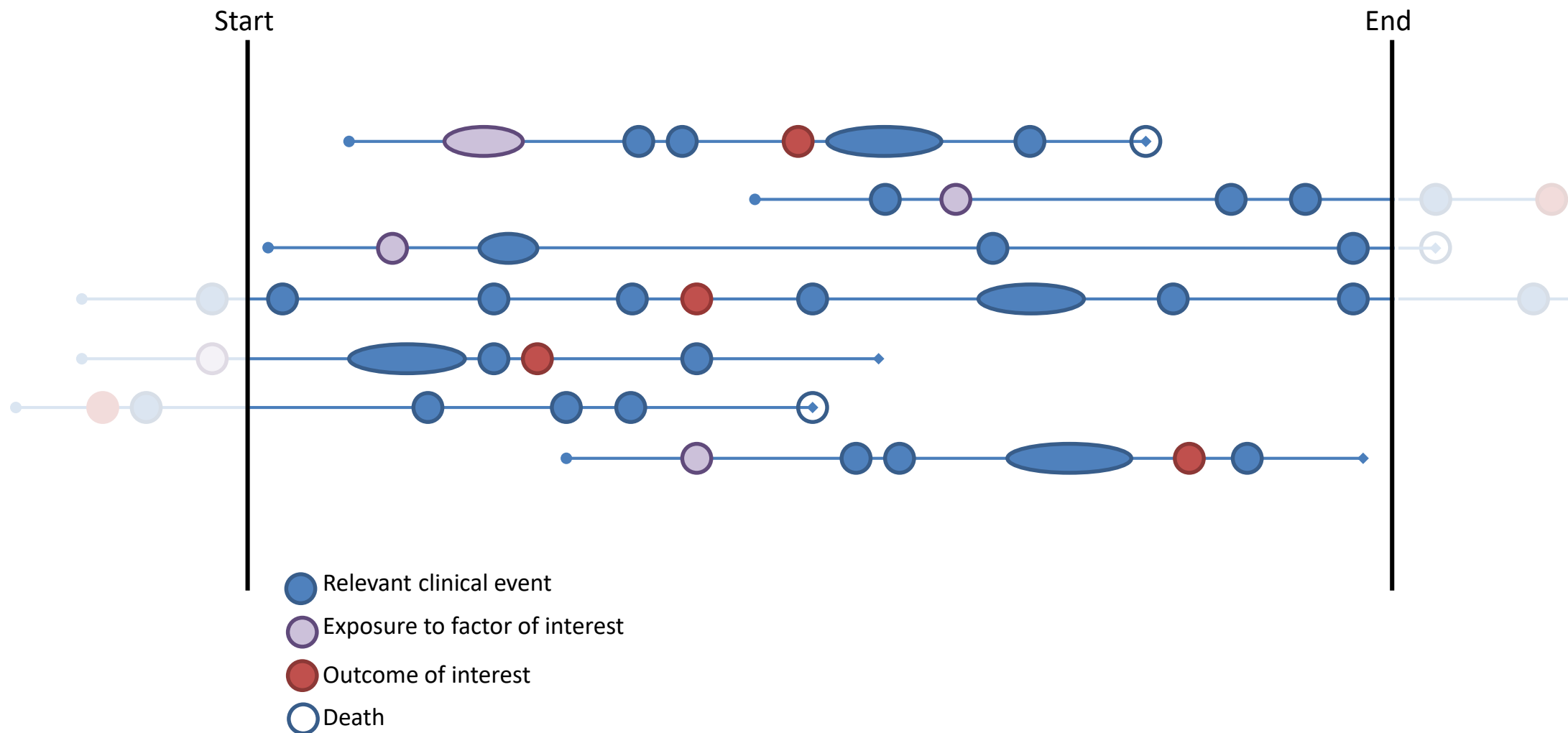Precise temporal allocation of Exposures and outcomes

Challenges
EHR's observation window

Opportunities

- Recent willingness by data custodians to research health data using machine learning based methodologies
- Wide range of exploratory or hypothesis generation/test studies
  - Patient classification (Machine Learning sub-phenotyping)
  - Detailed healthcare utilisation patterns (multi-state trajectory flows)
  - Integration of data models
  - Sophisticated epidemiological/statistical methods computationally feasible for causal inference
  - EHR based decision/early-detection tools (automation)

The Data Lab

Academic members

Spiros Denaxas
Senior Lecturer in Biomedical Informatics

Arturo Gonzalez-Izquierdo
Research Associate in Electronic Health Records

Kenan Direk
Research Associate in Electronic Health Records

Michalis Katsoulis
BHF Research Fellow

Marina Daskalopoulou
Research Associate

Ghazaleh Fatemifar
AHA Research Fellow

Natalie Fitzpatrick
Research Data Coordinator

Maria Pikoula
Clinical Data Scientist

Vaclav Papez
Clinical Data Scientist

Marcos Barreto
Royal Society Newton Fellow

Colin Josephson
Assistant Professor of Neurology and Community Health Sciences (Univ. of Calgary)

Marie Erwood
IMAGINE Data Manager

Natalie Fitzpatrick
Data Science Facilitator
n.fitzpatrick@ucl.ac.uk

CALIBER portal
https://www.caliberresearch.org/portal

Denaxas Lab
http://denaxaslab.org/