

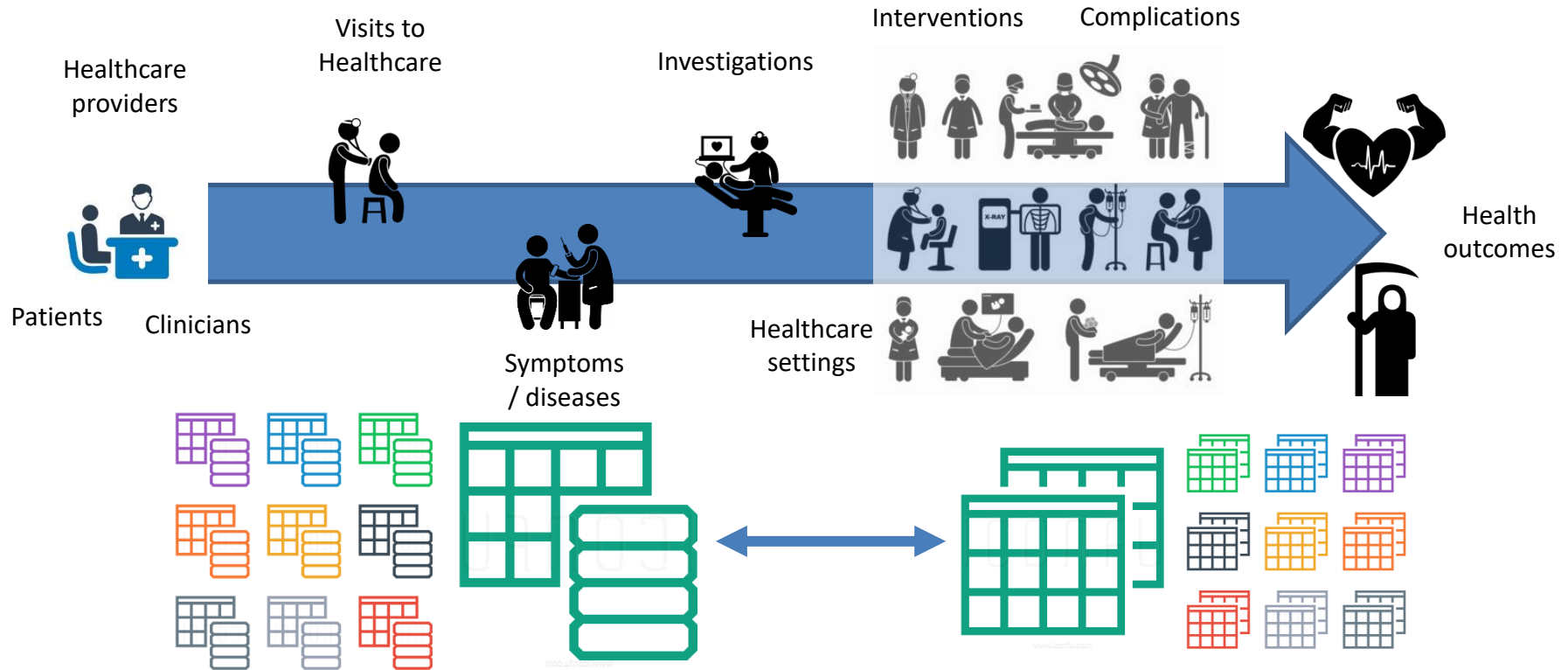
Introduction to Using
The Clinical Practice Research Datalink (CPRD)
**Phenotyping Electronic Health Records for
Research**

Arturo González-Izquierdo
Ania Zylbersztejn
Linda Wijlaars

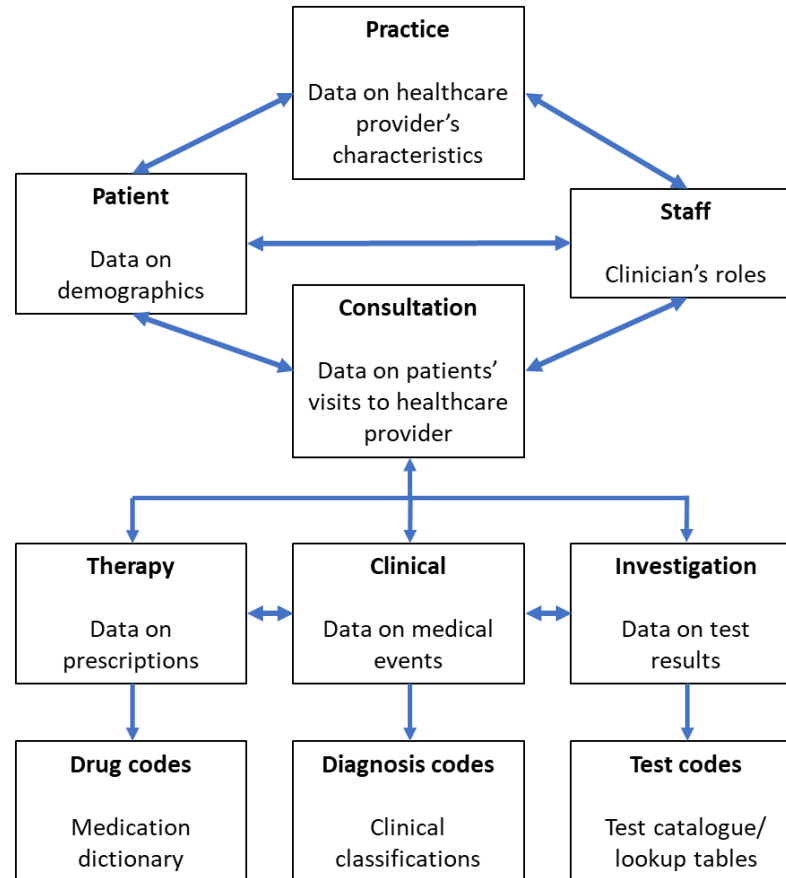
Objective

Provide the basic elements for the derivation of a phenotype using Electronic Health Records from Primary Care

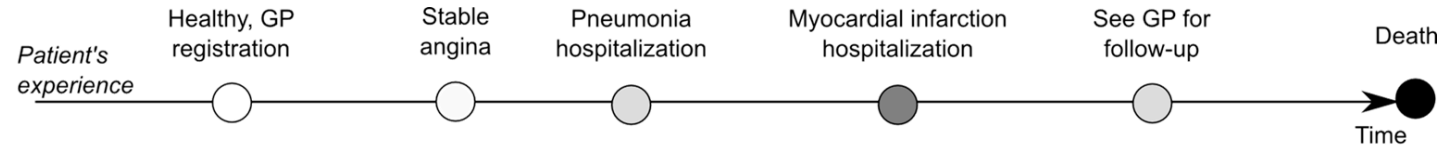
Data generation mechanism



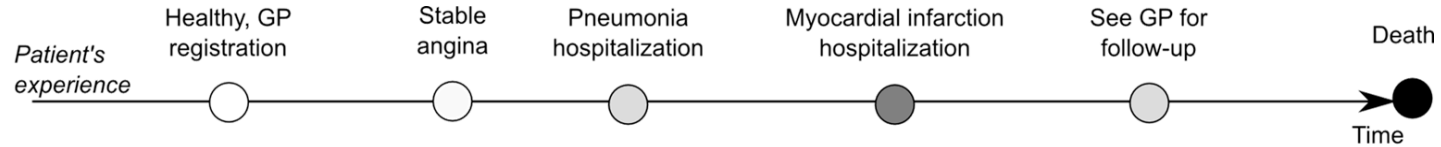
Data collection



Electronic Health Records

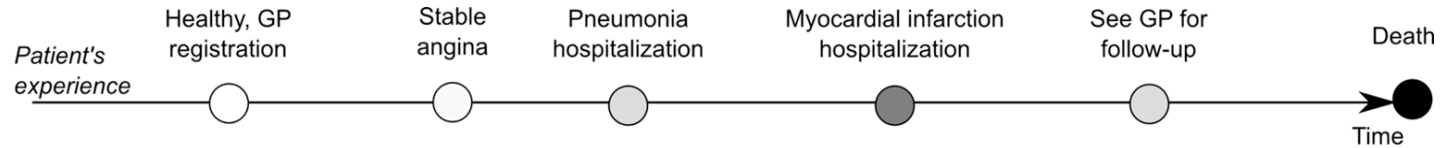


Electronic Health Records



Patient ID	Registration date	Consultation date	Admission date	Admission date	Consultation date	Date of death
	Date of birth	Diagnosis	Diagnosis	Diagnosis	Diagnosis	Underlying cause
	Gender	Blood pressure	Procedure	Procedure	Procedures	Subsidiary causes
	blood pressure	ECG reading	Discharge date	Prescribed treatment	Prescription	
	Weight	CT coronary angiography		ECG reading		
	Hight	Other tests		Cardiac biomarkers		
	Phisical activity	Prescribed treatment		Cardiac imaging		
	Health history (heart, diabetes, stroke)			Other tests		
	Blood pressure			Discharge date		
	Smoking, alcohol,...					
	contraception					
	immunisations					

Electronic Health Records



Patient ID	Registration date	Consultation date	Admission date	Admission date	Consultation date	Date of death
	Date of birth	Diagnosis	Diagnosis	Diagnosis	Diagnosis	Underlying cause
	Gender	Blood pressure	Procedure	Procedure	Procedures	Subsidiary causes
	blood pressure	ECG reading	Discharge date	Prescribed treatment	Prescription	
	Weight	CT coronary angiography		ECG reading		
	Hight	Other tests		Cardiac biomarkers		
	Phisical activity	Prescribed treatment		Cardiac imaging		
	Health history (heart, diabetes, stroke)			Other tests		
	Blood pressure			Discharge date		
	Smoking, alcohol,...					
	contraception					
	immunisations					

Electronic Health Records

- Medical classification – coding systems
 - Transformation of diagnoses, procedures and prescribed drugs into standard codes
 - For costing or resource planning purposes/ epidemiologic surveillance/ scientific research
 - Hierarchical structure
 - National or International

Electronic Health Records

- Coding systems in EHRs
 - Primary care:
 - Diagnosis, symptoms, signs... – Read or SNOMED codes
 - Prescribed medication – BNF, prodcodes
 - Secondary care
 - Diagnosis – ICD10
 - Procedures – OPCS

Electronic Health Records

Read clinical classification

- 0.... Occupations
- 1.... History / symptoms
- 2.... Examination / Signs
- 3.... Diagnostic procedures
- 4.... Laboratory procedures
- 5.... Radiology/physics in medicine
- 6.... Preventive procedures
- 7.... Operations, procedures, sites
- 8.... Other therapeutic procedures
- 9.... Administration
- A.... Infectious and parasitic diseases
- B.... Neoplasms
- C.... Endocrine, nutritional, metabolic and immunity
- D.... Diseases of blood and blood-forming organs
- E.... Mental disorders
- F.... Nervous system and sense organ diseases
- G.... Circulatory system diseases
- H.... Respiratory system diseases
- J.... Digestive system diseases
- K.... Genitourinary system diseases
- L.... Complications of pregnancy, childbirth and the
- M.... Skin and subcutaneous tissue diseases
- N.... Musculoskeletal and connective tissue disease
- P.... Congenital anomalies
- Q.... Perinatal conditions
- R.... [D]Symptoms, signs and ill-defined conditions
- S.... Injury and poisoning
- T.... Causes of injury and poisoning
- U.... [X]External causes of morbidity and mortality
- Z.... Unspecified conditions

Read Thesaurus	Read Code	Term Code	Term	
Occupations				
History / symptoms				
Depth of history	G30..	0	Acute myocardial infarction	Preferred Term
Family history				
Social/personal history	G30..	11	Attack – Heart	Synonym
Past medical history				
Gynaecological history	G30..	12	Coronary Thrombosis	Synonym
General symptoms				
Respiratory symptoms				
Cough	G30..	14	Heart attack	Synonym
No cough				
Dry cough	G30..	15	MI – Acute Myocardial Infarction	Synonym
Productive cough				
Productive cough -green sputum				
Productive cough-yellow sputum				
Productive cough NOS				
Night cough present				
Night cough absent				
Chesty cough				

Terms become more specific as you go down the hierarchy.

Electronic Health Records

SNOMED CT clinical classification (Systematized Nomenclature of Medicine Clinical Terms)

NHS Digital SNOMED CT Browser

© SNOMED International 2017 v1.36.4 - Hosted and maintained by NHS Digital

Summary Details Diagram Expression Refsets Members References Classification

Parents

- ≡ Congenital abnormality of cardiac connection (disorder)
- ≡ Congenital atrial septal defect (disorder)
- ≡ Congenital hypertrophy of cardiac ventricle (disorder)
- ≡ Congenital right ventricular anomaly (disorder)
- ≡ Congenital stenosis of pulmonary valve (disorder)
- ≡ Right ventricular hypertrophy (disorder)

≡ **Fallot's trilogy (disorder)** ☆

SCTID: 204351007

204351007 | Fallot's trilogy (disorder) |

- Fallot's trilogy
- Fallot's trilogy (disorder)
- Fallot triad
- Fallot trilogy

Occurrence → Congenital
Pathological process → Pathological developmental process
Finding site → Pulmonary valve structure
Associated morphology → Stenosis

Occurrence → Congenital
Associated morphology → Congenital septal defect
Finding site → Interatrial septum structure
Pathological process → Pathological developmental process

Occurrence → Congenital
Finding site → Entire right cardiac ventricle
Associated morphology → Hypertrophy
Pathological process → Pathological developmental process

Electronic Health Records

ICD-10 Version:2019

Search

▼ ICD-10 Version:2019

- ▶ I Certain infectious and parasitic diseases
- ▶ II Neoplasms
- ▶ III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- ▶ IV Endocrine, nutritional and metabolic disorders
- ▶ V Mental and behavioural disorders
- ▶ VI Diseases of the nervous system
- ▶ VII Diseases of the eye and adnexa
- ▶ VIII Diseases of the ear and mastoid process
- ▶ IX Diseases of the circulatory system
- ▶ X Diseases of the respiratory system
- ▶ XI Diseases of the digestive system
- ▶ XII Diseases of the skin and subcutaneous tissue
- ▶ XIII Diseases of the musculoskeletal system and connective tissue
- ▶ XIV Diseases of the genitourinary system
- ▶ XV Pregnancy, childbirth and the puerperium
- ▶ XVI Certain conditions originating in the perinatal period
- ▶ XVII Congenital malformations, deformations and chromosomal abnormalities
- ▶ XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified

Chapter V Mental and behavioural (F00-F99)

Incl.: disorders of psychological

Excl.: symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified

This chapter contains the

- [F00-F09](#) Organic, including alcohol-related
- [F10-F19](#) Mental and behavioural disorders due to use of alcohol
- [F20-F29](#) Schizophrenia, schizotypal and delusional disorders
- [F30-F39](#) Mood [affective] disorders
- [F40-F48](#) Neurotic, stress-related and somatoform disorders
- [F50-F59](#) Behavioural syndromes
- [F60-F69](#) Disorders of adult personality and behaviour
- [F70-F79](#) Mental retardation
- [F80-F89](#) Disorders of psychological development
- [F90-F98](#) Behavioural and emotional disorders with onset usually specific to childhood
- [F99-F99](#) Unspecified mental disorder

International Classification of Diseases (ICD10)

I61

Intracerebral haemorrhage

Excl.: sequelae of intracerebral haemorrhage ([I69.1](#))

I61.0

Intracerebral haemorrhage in hemisphere, subcortical
Deep intracerebral haemorrhage

I61.1

Intracerebral haemorrhage in hemisphere, cortical
Cerebral lobe haemorrhage
Superficial intracerebral haemorrhage

I61.2

Intracerebral haemorrhage in hemisphere, unspecified

I61.3

Intracerebral haemorrhage in brain stem

I61.4

Intracerebral haemorrhage in cerebellum

I61.5

Intracerebral haemorrhage, intraventricular

I61.6

Intracerebral haemorrhage, multiple localized

I61.8

Other intracerebral haemorrhage

I61.9

Intracerebral haemorrhage, unspecified

Electronic Health Records

patid	eventdate	readcode	data1	data2
1001	2011-12-01	246..00	67	67
1001	2012-06-05	G30..14		
1001	2012-06-05	8H2Z.00		
2221	2011-12-01	M211.12		
2221	2011-12-01	246..00	80	110
2221	2013-11-02	246..00	69	128
3241	2009-11-03	H26..00		
7810	2014-01-12	171..00		
7810	2014-01-12	H06z011		

Basic phenotyping

- How can we identify patients that meet the criteria of a specific medical condition, using routinely collected data?

Basic phenotyping

- Phenotype
- EHR Phenotype
 - Definition
 - Identification
 - Extraction
 - Use / Validation

Basic phenotyping - EHR phenotype

- Definition – what characteristic?
 - Clinical condition
 - Mental disorder
 - Physical trait
 - Biochemistry trait

Basic phenotyping - EHR phenotype

- Identification – what data are available and how is it generated?
 - Components (symptoms, diagnosis, treatment...)
 - Core vs supporting information
 - Precision (scales of measurement)

Basic phenotyping - EHR phenotype

- Definition – what characteristic?
 - Clinical condition: **Hypertension**

A patient will be classified as having hypertension if their health records show: A diagnosis of hypertension during a GP consultation or during a hospitalisation; or blood pressure readings: at least 2 systolic or diastolic high blood pressure readings within a 6 month period (systolic_BP \geq 150 mmHg OR diastolic_BP \geq 90 mmHg). For diabetic patients and patients over 60 years old, a lower threshold is used (SYS_BP \geq 140 mmHg OR DIAS_BP \geq 90 mmHg); or at least two prescriptions of blood pressure lowering medication within 6 months

Basic phenotyping - EHR phenotype

- Extraction – How will we derive the new data?
 - Identify specific fields from each data source
 - Define specific instructions to transform the available information into components of the phenotype
 - Define the level of granularity to express the phenotype
 - Systematic across individuals – Phenotype Algorithm!

Basic phenotyping - EHR phenotype

IF HT = 1, THEN HT_comp = 1

Biometrics, test results, time dependent thresholds

For non-diabetic patients:

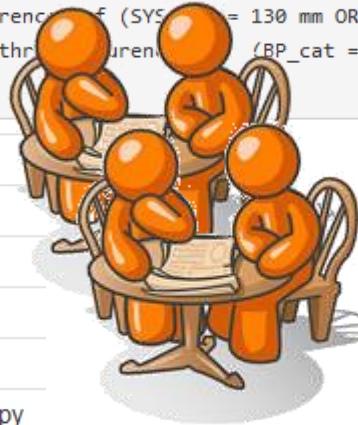
IF there are at least three occurrences of (SYS_BP >= 140 mmHg OR DIAS_BP >= 90 mmHg) within 1 year of relevant date) OR if there are at least three occurrences of (BP_cat = 3) within 1 year of relevant date
THEN HT_comp = 1

For diabetic patients:

IF there are at least three occurrences of (SYS_BP >= 130 mm OR DIAS_BP >= 80 mmHg) within one year of relevant date) OR if there are at least three occurrences of (BP_cat = 3) within 1 year of relevant date
THEN HT_comp = 1

Diagnoses

6627.00	Goal hypertension
6628.00	Possible hypertension
662F.00	Hypertension
662G.00	Hypertensive treatment 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
7001.00	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

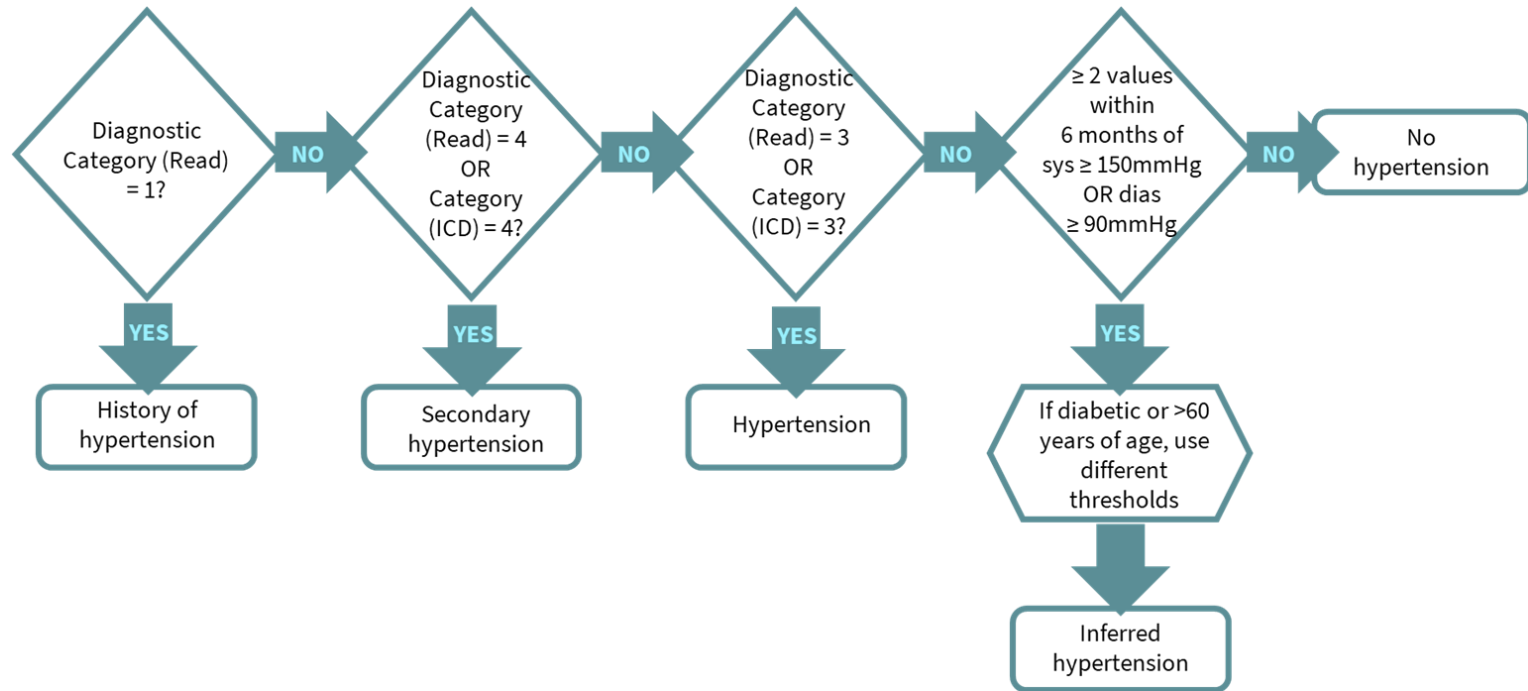
Health care utilisation patterns

Basic phenotyping - EHR phenotype

- Developing a new code list
- HDR UK Phenotyping Library

<https://phenotypes.healthdatagateway.org/>

Basic phenotyping - EHR phenotype



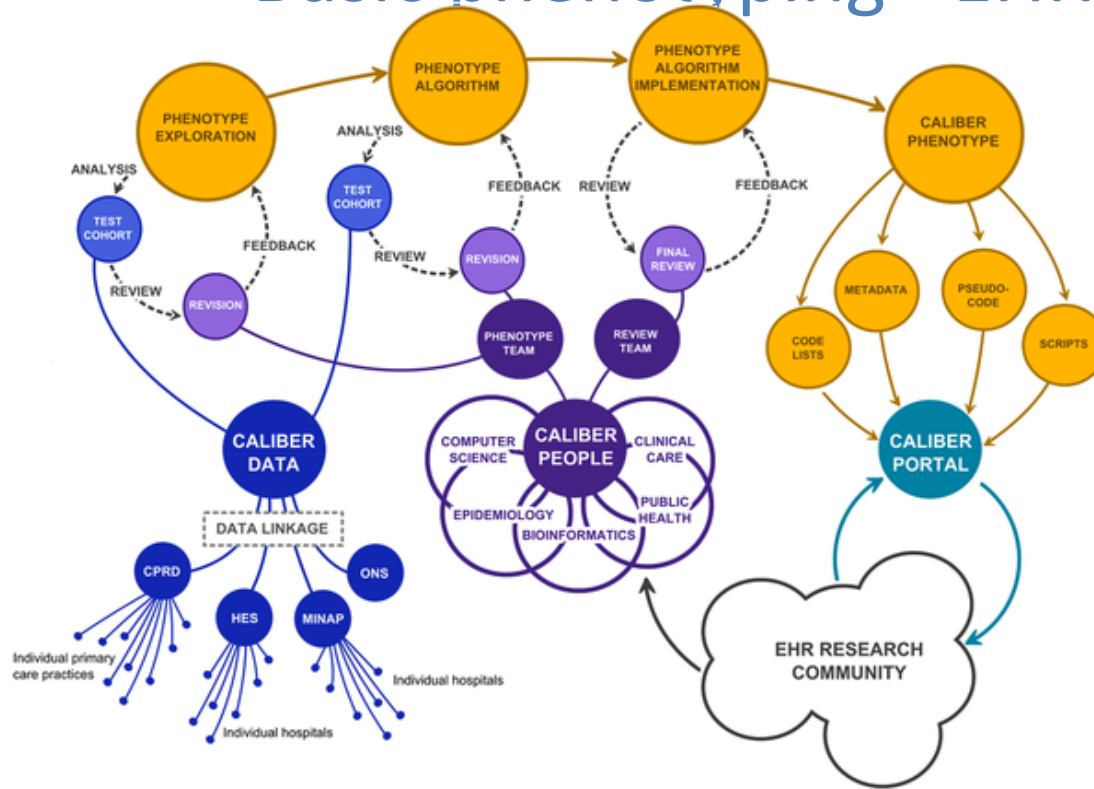
Basic phenotyping - EHR phenotype

Increasing phenotype granularity according to available data			
Binary (presence/ absence)	Multiple categories		In depth
	Certainty	Severity	Propensity score
Yes	“Definite”	Severe	Continuous scale
		Moderate	
		Mild	
	Probable	Probable	
	Possible	Possible	
No	No	No	

Basic phenotyping - EHR phenotype

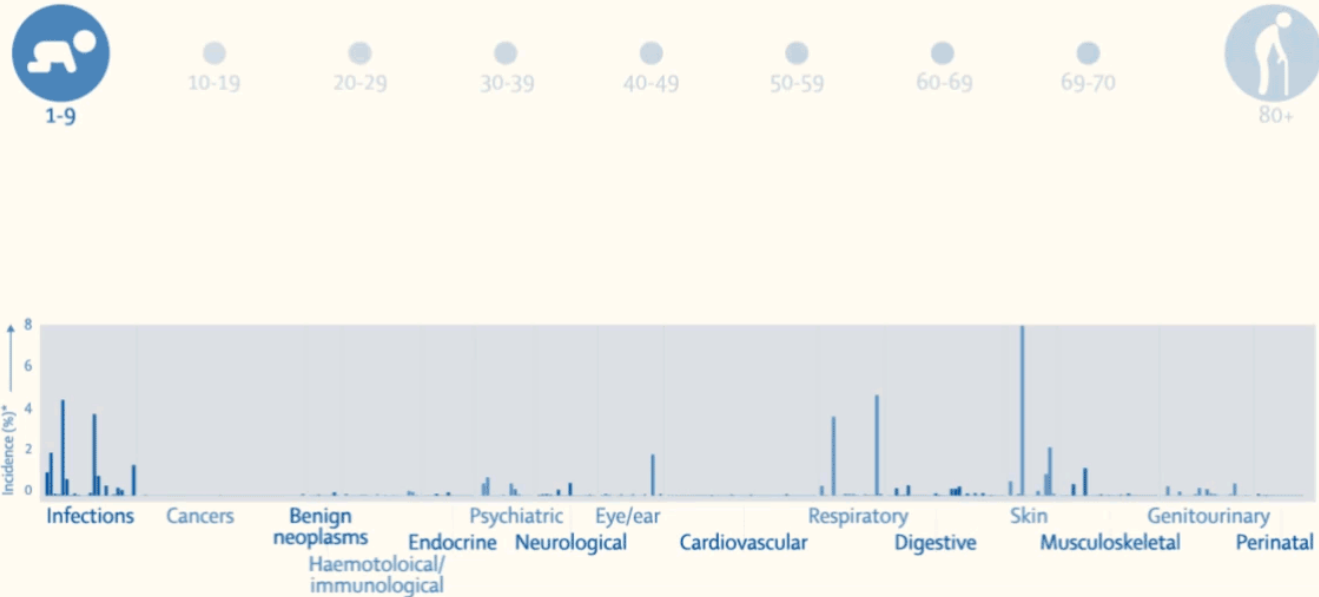
Hypertension phenotype		
Binary classification	Multiple categories	
	Certainty	Classification
Yes	Diagnosed Hypertension	History of hypertension
		Hypertension
		Secondary hypertension
	Inferred hypertension	Inferred hypertension
No	No	No

Basic phenotyping - EHR phenotype



Chronological map of human disease

How health changes over life



THE LANCET Digital Health

The best science for better lives

V. Kuan et al. The Lancet Digital Health 2019

NCRM Research Methods e-Festival - November 2023