A Prognostic Atlas of Clinical Medicine: some initial steps

Berlin Institute of Health Lecture Series
'Frontiers in Translational Medicine – Scientific and Structural Challenges'

26 November 2021

Harry Hemingway
What do doctors do? What do hospitals do?
Doctors treat patients and they keep records > EHR data

*In-hospital transfers* 2.4 m patients (8m transfers)
over 10 years
Clinicians diagnose diseases and write it down

56 million people in England alive at 30 Jan 2020 (the day before COVID case #1) and updated hospitalisations, currently 250 million

Prof Cathie Sudlow OBE
Prof Spiros Denaxas
Healthcare systems across the world know remarkably little about which patients have which diseases, in which combinations, and with what outcomes.
International unmet need
Lack of findable, accessible, useful personalised risk for people with each disease

'People like me'

Existing risk information:

• Low coverage of: <5% of diseases have a prognostic model

• Not 'for' patients

'Patients like mine'

• Mono morbid approach (but patients commonly multimorbid)

• Serendipitous ‘one disease at a time’ generation of prognostic models: no systematic framework
"Patients have the right to discuss risks"

"You must ... share information on likely progression"

Can we be more systematic and do this for each and every disease?

Prognosis and medical ethics
Democratic approaches to answering ubiquitous questions across the ‘long tail’ of clinical medicine

Unique diseases managed in a health system (ranked by frequency)

- All specialities
- All clinicians
- All diseases
- All people with diseases
Prognosis: massive need for more, and more actionable, research

**Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes.**

**Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research.**

**Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research.**

**Prognosis Research Strategy (PROGRESS) 4: Stratified Medicine Research.**
Approach: from serendipity to systems
One disease at a time
1 Disease: Angiographic Coronary Artery disease:
Clinical data: 5 hospitals, ‘weighing paper case notes’
Outcomes from hospital EHR linkage (first time in England)

Hemingway et al. NEJM 2001
1 Disease: angina
Clinical data: whole country nationwide in Finland
CD sent in the post!

<table>
<thead>
<tr>
<th>Age Group, y</th>
<th>Sex</th>
<th>Observed/ Expected Deaths</th>
<th>SMR (95% CI)</th>
<th>P Value for Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>45-54</td>
<td>Women</td>
<td>11/1</td>
<td>12.1 (6.06-21.7)</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>120/21</td>
<td>5.63 (4.71-6.74)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>Women</td>
<td>55/12</td>
<td>4.69 (3.60-6.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>229/95</td>
<td>2.40 (2.11-2.73)</td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>Women</td>
<td>221/88</td>
<td>2.50 (2.20-2.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>452/242</td>
<td>1.87 (1.70-2.05)</td>
<td></td>
</tr>
<tr>
<td>75-84</td>
<td>Women</td>
<td>322/187</td>
<td>1.72 (1.54-1.92)</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>329/188</td>
<td>1.75 (1.57-1.95)</td>
<td></td>
</tr>
<tr>
<td>85-89</td>
<td>Women</td>
<td>127/64</td>
<td>2.00 (1.68-2.37)</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>74/39</td>
<td>1.93 (1.54-2.42)</td>
<td></td>
</tr>
</tbody>
</table>
1 Disease: acute myocardial infarction

Whole country nationwide in UK and Sweden

Disease registry: manual data entry, not part of EHR

---

**Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK**

Sheng-Chieh Chang, Rolf Gotlob, Owen Nicholas, Steven James, Anders Jeppsson, Charles Wolfe, Peter Hauschmann, Lars Wallentin, John Deanfield, Adam Tibbis, Tomas Jeromy, Harry Hemingway
12 diseases at a time
12 diseases: Higher resolution ‘CVD’

3% sample of England (1.2m people)
EHR (primary care linked to secondary care)
clinically recorded blood pressure

Informed American and European guidelines

Risks differ across diseases

Rapsomaniki et al. The Lancet 2014
Many common diseases
A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service

Valerie Kuan, Spinoz Demouza, Arturo Gonzalez-Lagueda, Kieran Dirk, Osman Bhattach, Shamez Hussain, Shaile Sutaria, Melanie Hingpranee, Dorothea Nitsch, Constantinos A Parisinos, R Thomas Lumbiers, Rohini Mathur, Rechoa Sofiat, Juan P Cosas, Ian C Wong, Harry Hemingway, Arnon D Hingpranee

How health changes over life

How health changes over life

THE LANCET

Dr Valerie Kuan
EHR phenotyping

1001, 2000-01-01, 23,1,NULL,H48
1121, 2013-06-04, 7,1,3,5,14AN,00
1121, 2011-06-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1,00
1511, 199-03-11, 91,1,6, G573100
9913, 2012-05-21, 81,1,9, G573100
67222, 1994-11-01, 1234,1,3,7L1H300
1001, 1994-08-11, 1234,1,3,7L1H300
1001, 1993-01-01, 253,1,1,793Mz60
1231, 2012-03-03, 23,1,123,K65
1121, 2013-05-04, 7,1,3,5,14AN,00
1121, 2011-05-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1,00
1511, 199-03-11, 91,1,6, G573100
9913, 2012-05-21, 81,1,9, G573100
67222, 1994-11-01, 1234,1,3,7L1H300
67222, 1995-12-21, 1234,1,3,7L1H300
67222, 1991-03-03, 1234,1,3,7L1H300
68244, 1993-01-01, 253,1,1,793Mz60
1121, 2013-05-04, 7,1,3,5,14AN,00
1121, 2011-05-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1,00
1511, 199-03-11, 91,1,6, G573100
67222, 1995-12-21, 1234,1,3,7L1H300
67222, 1994-11-01, 1234,1,3,7L1H300

Disease status
Severity
Onset
Certainty

EHR Phenotyping
Open EHR Phenotyping library

The HDR UK Phenotype Library is a comprehensive, open access resource providing the research community with information, tools and phenotyping algorithms for UK electronic health records.

Search our Phenotype library

- 753 Phenotypes
- 1462 Concepts
- 101562 Clinical Codes
- 22 Data Sources
- 14 Coding Systems

A Reference Catalogue of Human Diseases

Connected. The Phenotype Library is accessible via an API to support interoperability, is integrated with health dataset information in HDR-UK’s Innovation Gateway, and hosts content from numerous contributing organisations.

Patient-focused. The Library is enabling important research to improve patient health and well-being. Content spans major disease areas, including heart disease, cancer, COVID-19 and other common and rare human health conditions. Curated collections from contributors such as the HDR UK BREATHE Hub for respiratory health share clinical expertise to tackle critical research questions.

Cutting-edge. Built with a focus on computability, this resource aims to drive the next generation of research methods. Integration with Phenoflow enables executable implementations of the phenotypes in our collection, while the API and R package client facilitate integration of the Library content directly into other analysis workflows.

https://phenotypes.healthdatagateway.org
Asthma
Eleanor L Axson, Jennifer K Quint

Type: Disease or Syndrome
ID: PH12
Version ID: 24

Data Sources:
- Clinical Practice Research Datalink GOLD
- Clinical Practice Research Datalink AURUM
- Hospital Episode Statistics APC for CPRD GOLD
- Hospital Episode Statistics APC for CPRD Aurum
- Death Registration data for CPRD GOLD
- Death Registration data for CPRD Aurum
- UK Biobank

Valid event data range: 01/01/2001 - 31/12/2019
Sex: Female, Male
Agreement Date: 2020-06-03
Coding Date: Read codes v2, ICD10 codes, SNOMED codes, UKBiobank codes, ICD11 codes

Tags: BREATHE, Phenotype Library

Definition

These codes will capture asthma ever, not just current asthma. These codes are not intended to be mandatory, but are to be used as a starting point for the identification of asthma in routine EHR. Each study may differ in the sensitivity and specificity of the coding required.

For those interested in further discrimination of asthma phenotypes, we refer you to Nissen et al. 2019.


Validation of Read Codes for the Identification of COPD in CPRD
Quint et al. validated a set of Read codes for the identification of COPD in CPRD in 2014. Using diagnostic codes alone, the positive predictive value (PPV) was 86.5% (77.592.3%). Requiring a diagnostic code, spirometry measures, and specific medication increased PPV to 89.4% (80.7794.5%) but reduced case numbers by 10%.

https://phenotypes.healthdatagateway.org
Validations of EHR Phenotypes

- Specialist adjudication vs standard
- Clinical relevance e.g. mapping to guidelines + quality initiatives
- Outcomes and Prognosis
- Genetic / molecular / aetiology
- Concordance across settings (1ry, 2ry, 3ry)
- Transportability across health systems and nations
Uses and insights from EHR Phenotypes

- Specialist adjudication vs standard
- Clinical relevance e.g. mapping to guidelines + quality initiatives
- Outcomes and Prognosis
- Genetic / molecular / aetiology
- Concordance across settings (1ry, 2ry, 3ry)
- Transportability across health systems and nations

Denaxas et al. JAMIA 2018
Use of 100s of EHR disease phenotypes: pre-prototype of prognostic atlas for pandemic response

10 HOURS AGO by Clive Cookson in London

The latest UK government strategy to slow the coronavirus epidemic could lead to between 35,000 and 70,000 excess deaths over the next year, according to an instant analysis by scientists from University College London, the University of Cambridge and Health Data Research UK.

Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study

Amitava Banerjee, Laura Pesse, Steve Harris, Arturo González-Izquierdo, Ana Terral-Bo, Laura Shallcross, Mahshid Nour sadeghi, Deenesh Pillay, Neil Shah, Chris Holmes, Christine Pajdl, Wei Keong Wong, Claudia Langenberg, Bryan Williams, Spinos Denisac, Harry Hemingway

Pre-print 21 March 2021
The day before UK Lockdown 1
“At last I have some information about me that I can act on”
1.4m page views
Use of 100s of EHR disease phenotypes: pre-prototype of prognostic atlas lockdown 1 - cancer

Atlas as ‘canary in the mine’

Referrals for urgent cancer diagnosis
subsequent NHS Recovery Plan in July 2020 prioritizes cancer services
2021 - a step up in the scale and depth of data in England accessible by researchers

Prof Cathie Sudlow OBE

Dr Johan Thygesen

Dr Chris Tomlinson
COVID trajectories in 56m people

COVID-19 Mortality - Stratified by worst healthcare presentation

A) Wave 1

B) Wave 2

Cumulative number of events

Number at risk

Cumulative number of events

Positive test - 62624
Primary care diagnosis - 94011
Hospitalisation - 83332
ICU admission - 10671
Critical care outside ICU - 5282

Positive test - 797707
Primary care diagnosis - 1845299
Hospitalisation - 194284
ICU admission - 23747
Critical care outside ICU - 15003

0 7 14 21 28
Time

0 20 40 60
Cumulative mortality (%)

0 20 40 60
Cumulative mortality (%)

Tygesen et al. medRxiv (preprint) 9 November 2021
All diseases across common-rare disease continuum that leave a digital trace
Clinicians diagnose diseases and write it down
56 million people in England alive at 30 Jan 2020 (the day before COVID case #1) and updated hospitalisations, currently 250 million

N Patients (log)

ICD-10 codes

- Allergy & Immunology - ALL
- Cardiac Electrophysiology – CARDEL
- Cardiac Preventative - CARDPR
- Cardiology - CARD
- Dermatology - DERM
- Endocrinology - ENDO
- Gastroenterology - GAST
- General Practice - GENPRA
- Genetics - GEN
- Gynaecology - GYN
- Gynaecology Oncology - GYNONCO
- Haematology - HAEM
- Haematology Oncology - HAEMONCO
- Hepatology - HEPA
- Infectious & Tropical - INF
- Neonatology - NEO
- Neurology - NEUR
- Neurosurgery - NEURSUR
- Obstetrics - OBS
- Oncology - ONCO
- Oral - ORAL
- Ophthalmology - OPTH
- Orthopaedics - ORTHO
- Otalaryngology - ORL
- Paediatrics - PEAD
- Paediatric Cardiology - PAEDCARD
- Psychiatry - PSYCH
- Public Health - PUB
- Renal - REN
- Respiratory - RESP
- Rheumatology - RHEU
- Stroke Medicine - STRO
- Sexual Health - SEXU
- Surgery - SUR
- Urology - URO
- Vascular Surgery - VASCSUR
• It is the *most* widely used terminology internationally in health systems, and research (morbidity and mortality)

• 22 chapters (based on pathology, or physiology or anatomy)

• 12000 of 17 000 unique ICD-10 4 character codes *are used in practice*

• Hi fidelity across common and rare disease abundance

• Does not readily Classify diseases!
  - 1 disease may have many leaf codes in same chapter: clinical role in grouping e.g. Vanderbilt PheCodes
  - 1 disease may have codes ‘fragmented’ across multiple chapters.
  - 1 code can map to many diseases
  - Does not distinguish disease from non–disease codes (e.g abnormal tests)

• Does not readily engage Clinical Specialists (who create data in first place)
D-code: from ‘ICD chapter’ to clinical specialty

Chapter codes (3 characters)

A00–B99
C00–D48
D50–D89
E00–E90
F00–F99
G00–G99
H00–H59
H60–H95
I00–I99
J00–J99
K00–K93
L00–L99
M00–M99
N00–N99
O00–O99
P00–P96
Q00–Q99
R00–R99
S00–T98
V01–Y98
Z00–Z99
U00–U99

Chapter

Certain infectious and parasitic diseases
Neoplasms
Diseases of the blood and blood-forming organs
Endocrine, nutritional and metabolic diseases
Mental and behavioural disorders
Diseases of the nervous system
Diseases of the eye and adnexa
Diseases of the ear and mastoid process
Diseases of the circulatory system
Diseases of the respiratory system
Diseases of the digestive system
Diseases of the skin and subcutaneous tissue
Diseases of the musculoskeletal system and connective tissue
Diseases of the genitourinary system
Pregnancy, childbirth and the puerperium
Certain conditions originating in the perinatal period
Congenital malformations, deformations and chromosomal abnormalities
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
Injury, poisoning and certain other consequences of external causes
External causes of morbidity and mortality
Factors influencing health status and contact with health services
Codes for special purposes

Clinical specialties:
- Allergy & Immunology
- Cardiac Electrophysiology
- Cardiac Paediatrics
- Cardiac Preventative
- Cardiology
- Dermatology
- Endocrinology
- Gastroenterology
- General Practice
- Genetic & Metabolomic Medicine
- Gynaecology
- Gynaecology Oncology
- Haematology
- Haematology Oncology
- Hepatology
- Infectious & Tropical Medicine
- Lymphology
- Neonatology
- Neurology
- Neurosurgery
- Obstetrics
- Oncology
- Ophthalmology
- Oral
- Orthopaedics
- Otorhinolaryngology
- Paediatrics
- Psychiatry
- Public Health
- Renal
- Respiratory
- Rheumatology
- Stroke Medicine
- Sexual Health
- Surgery
- Urology
- Vascular Surgery
D-code: from ‘ICD chapter’ to clinical specialty

From 12k 4 character ICD-10 codes

Using Vanderbilt PheCodes (1300 leaf codes), we are developing data informed D codes: pre-beta version.

5390 unique "Disease" type phecodes

Of which
3244 are leaf codes

Of which
82 are "other specified"

Of which
43 offer specifications in rubric
Prognostic Atlas ⇐ Map to therapeutic areas of top 10 pharma companies

Pre-competitive opportunities to sponsor knowledge generation / management, across conditions in a common framework

Source: Pharmaprocess®️, January 2020
Cardiac dysrhythmias

- Paroxysmal tachycardia, unspecified (I41.8) (Cardiology-Electrophysiology)
- Paroxysmal supraventricular tachycardia (I43.2) (Cardiology-Electrophysiology)
- Paroxysmal ventricular tachycardia (I43.4) (Cardiology-Electrophysiology)
- Atrial fibrillation and flutter (I48) (Orphanet 1) (Cardiology-Electrophysiology)
- Atrial fibrillation (Cardiology-Electrophysiology)
- Atrial flutter (Cardiology-Electrophysiology)
- Paroxysmal atrial fibrillation (I48.9) (Orphanet 1) (Cardiology-Electrophysiology)
- Other specified cardiac dysrhythmias (I49) (Cardiology-Electrophysiology)
- Other specified cardiac arrhythmias (I49.8) (Cardiology-Electrophysiology)

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>ICD Label</th>
<th>ICD rubric</th>
<th>N people with Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>I49.8</td>
<td>Other specified cardiac arrhythmias</td>
<td>Brugada syndrome Long QT syndrome Rhythm disorder: coronary sinus, ectopic, nodal</td>
<td>59,483</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OrphaCode</th>
<th>Orphanet Label</th>
<th>Map type</th>
</tr>
</thead>
<tbody>
<tr>
<td>8022</td>
<td>Brugada Syndrome</td>
<td>Exact</td>
</tr>
<tr>
<td>10671</td>
<td>Torsade-de-pointes syndrome with short coupling interval</td>
<td>Narrow-Broad</td>
</tr>
<tr>
<td>10670</td>
<td>Familial short QT syndrome</td>
<td>Narrow-Broad</td>
</tr>
<tr>
<td>28037</td>
<td>GNB5-related intellectual disability-cardiac arrhythmia syndrome</td>
<td>Narrow-Broad</td>
</tr>
</tbody>
</table>
Importance of growing use of SNOMED CT

• ‘Common language’ across primary and secondary and tertiary care

• Semantic

• Higher clinical resolution, with >300k terms

• So may add considerable value in building a reference catalogue of disease, and prognostic atlas

........................................but similar need for Clinical Speciality engagement
D-code clinical speciality ownership, authoring and review: SNOMED-CT

Key to buttons for each concept
- Expand  Show descendants of this concept
- Contract  Hide descendants of this concept
- ?  Mark as unsure (to check)
-  Add a concept
- - Remove a concept
- ++ Add a concept and all descendants
- -- Remove a concept and all descendants

Reviewing tools
- Show top-level concepts only
- Show all concepts
- Mark all concepts as "checked"
- Mark all concepts as "unchecked"
- Show unchecked concepts only

Export to Heart failure hierarchy .csv

<table>
<thead>
<tr>
<th>Expand</th>
<th>SNOMED CT concept</th>
<th>Comment</th>
<th>Checked</th>
<th>Included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heart failure with reduced ejection fraction (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
<tr>
<td></td>
<td>Heart failure with reduced ejection fraction due to cardiomyopathy (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
<tr>
<td></td>
<td>Heart failure with reduced ejection fraction due to coronary artery disease (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
<tr>
<td></td>
<td>Heart failure with reduced ejection fraction due to heart valve disease (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
<tr>
<td></td>
<td>Heart failure with reduced ejection fraction due to myocarditis (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
<tr>
<td></td>
<td>High output heart failure (disorder)</td>
<td>Add...</td>
<td>Y</td>
<td>? ++-+-+</td>
</tr>
</tbody>
</table>
D-code: operationalising the ‘treatable diseasome’

- Is the disease the subject of an evidence based clinical practice guideline?

Or

- Is the disease treated with an orphan medication approved by FDA, EMA, MHRA?

~700 distinct diseases
D-code > draft a clinically useful reference catalogue of disease (to help answer question: what do doctors do, what do hospitals do?)

Raw ICD-10
22 Chapters: heterogeneous

D-Code annotation in the light of all data
>> 45 clinical specialities, with review tools

12k Leaf codes used: heterogeneous

>>~3,000 disease ‘leaf’ codes for clinician review, with data for QC and engagement. If and where relative abundance makes sense, what are speciality driven priority uses?

>> Clinical prioritisation of ~700 diseases
  mapping to clinical practice guideline or
  An approved orphan medication
Prognostic atlas: identifying and connecting underpinning elements

Diseases
Data
Doctors
Guideline generating bodies
Charities

Dr Arturo Gonzalez-Izquierdo
Muhammad Qummer ul Arfeen
Dr Evaleen Malgapo
Natalie Fitzpatrick
Dr Serina Hayes
John Dinnwell
Izzie Harvey
Guideline generating bodies
Named doctors, their speciality and (often) their sub-sub speciality.
Patient organisations, charities
Often disease specific
Prognostic Atlas prototype

Specialities as ‘chromosomes’ (outermost track)

Small subset of conditions and specialities for display purposes
Primary pulmonary hypertension

36 000 people in England

Prevalence 6.4 per 10 000 (exceeding rare disease threshold)
Cross speciality Multimorbidity: Example of common cancers
<table>
<thead>
<tr>
<th>Phecode</th>
<th>N deaths</th>
<th>N cases Phecode</th>
<th>N covid19</th>
<th>% covid19</th>
<th>HR (95% CI)</th>
<th>Crude Prevalence (10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embolism and thrombosis of renal vein</td>
<td>354</td>
<td>2533</td>
<td>300</td>
<td>11.8</td>
<td>1.97 (1.47-2.65)</td>
<td>44.74</td>
</tr>
<tr>
<td>Intracerebral haemorrhage in brain stem</td>
<td>441</td>
<td>2441</td>
<td>272</td>
<td>11.14</td>
<td>1.98 (1.51-2.60)</td>
<td>43.12</td>
</tr>
<tr>
<td>Embolism and thrombosis of vena cava</td>
<td>1044</td>
<td>6470</td>
<td>733</td>
<td>11.32</td>
<td>2.05 (1.72-2.45)</td>
<td>114.29</td>
</tr>
<tr>
<td>Polyarteritis nodosa</td>
<td>230</td>
<td>3606</td>
<td>367</td>
<td>10.1</td>
<td>8.01 (6.09-10.5)</td>
<td>63.70</td>
</tr>
<tr>
<td>Wegener's granulomatosis</td>
<td>569</td>
<td>10490</td>
<td>918</td>
<td>8.7</td>
<td>8.58 (7.16-10.28)</td>
<td>185.30</td>
</tr>
<tr>
<td>Vulval varices</td>
<td>21</td>
<td>1636</td>
<td>136</td>
<td>8.3</td>
<td>9.28 (3.33-25)</td>
<td>28.39</td>
</tr>
</tbody>
</table>

HR (95%CI) adjusted by Age, Sex and Ethnicity
143 phecodes with N covid cases >100 (Illustration purpose)
Frequency map of disease: cardiology

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>N patients in NHS</th>
<th>Example diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in 10</td>
<td>5 000 000</td>
<td>Hypertension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>atrial fibrillation</td>
</tr>
<tr>
<td>1 in 100</td>
<td>500 000</td>
<td>Aortic stenosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Endocarditis</td>
</tr>
<tr>
<td>1 in 1000</td>
<td>50 000</td>
<td>Primary pulmonary hypertension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypertrophic cardiomyopathy</td>
</tr>
<tr>
<td>1 in 10 000</td>
<td>5 000</td>
<td>Endocardial myofibroelastosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Takayasu's disease</td>
</tr>
<tr>
<td>1 in 100 000</td>
<td>500</td>
<td>Ruptured papillary muscle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kyphoscoliotic heart disease</td>
</tr>
<tr>
<td>1 in 1 000 000</td>
<td>50</td>
<td>Endomyocardial eosinophilia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Candidal endocarditis</td>
</tr>
</tbody>
</table>

5 per 10 000 threshold

‘Common’

‘Rare’
Usefulness of existing prognostic models: from literature, one disease at a time

- Available for <5% of diseases
- Proliferate for some diseases e.g. >50 for heart failure
- Clinical practice guidelines rarely engage in prognosis, prognostic models
- In practice clinicians commonly report being ‘in the dark’
- Opportunity for engineered prognostic models across diseases of common form
  - Near term: Age, index condition, co-existing conditions
  - Longer term: incremental prognostic value (for a given *purpose*) of molecular and other information
Identifying Drug Targets and Disease Mediators

RESERCH ARTICLE SUMMARY

DISEASE GENOMICS

Mapping the proteo-genomic convergence of human diseases

Maik Pietzner†, Eleanor Wheeler†, Julia Carrasco-Zanini, Adrian Cortes, Mine Kopru, Maria A. Wörheide, Erin Oerton, James Cook, Isobel D. Stewart, Nicola D. Kerrison, Jian’an Luan, Johannes Raffler, Matthias Arnold, Wiebke Arit, Stephen O’Rahilly, Gabi Kastenmüller, Eric R. Gamazon, Aroon D. Hingorani, Robert A. Scott, Nicholas J. Wareham, Claudia Langenberg®

1,859 connections

Proteo-genomic map of human health

www.omicscience.org

412 protein targets and 506 curated phenotypes

Harmonization across EHR resources in the UK Biobank

N~1,500 phecodes ordered by frequency

Pietzner et al. Science 2021
Potential uses of atlas, driven by patients and specialists

• ‘How has the pandemic affected my chances of survival’

• Need for health systems to embed ‘Canaries in the mine’

• Shared decision making

• Clinical audit / quality of care / quality of data ‘Keogh principle’

• Trial feasibility and design

• …………..ingenuity
Conclusion

A prognostic atlas across clinical medicine which is specialist and patient driven has become feasible in the light of current data opportunities.

If it is a duty of doctors, and a right of patients, to understand the likely course of disease, then this may be a responsibility.
Thank you