Institute
and Faculty
of Actuaries

Big health data: perspectives across the patient journey from linking multiple record sources

Harry Hemingway FFPH, FRCP

Professor of Clinical Epidemiology

Director, Farr Institute of Health Informatics Research, UCL

15-17 September 2014, Birmingham

Outline

- What are big health data?
- Why me – personal journey
- What are big data good for?
  - Discovery
  - Trials
  - Outcomes, risk prediction and clinical decision making
  - Public health
- What is the role of the Farr Institute?

What are big data?

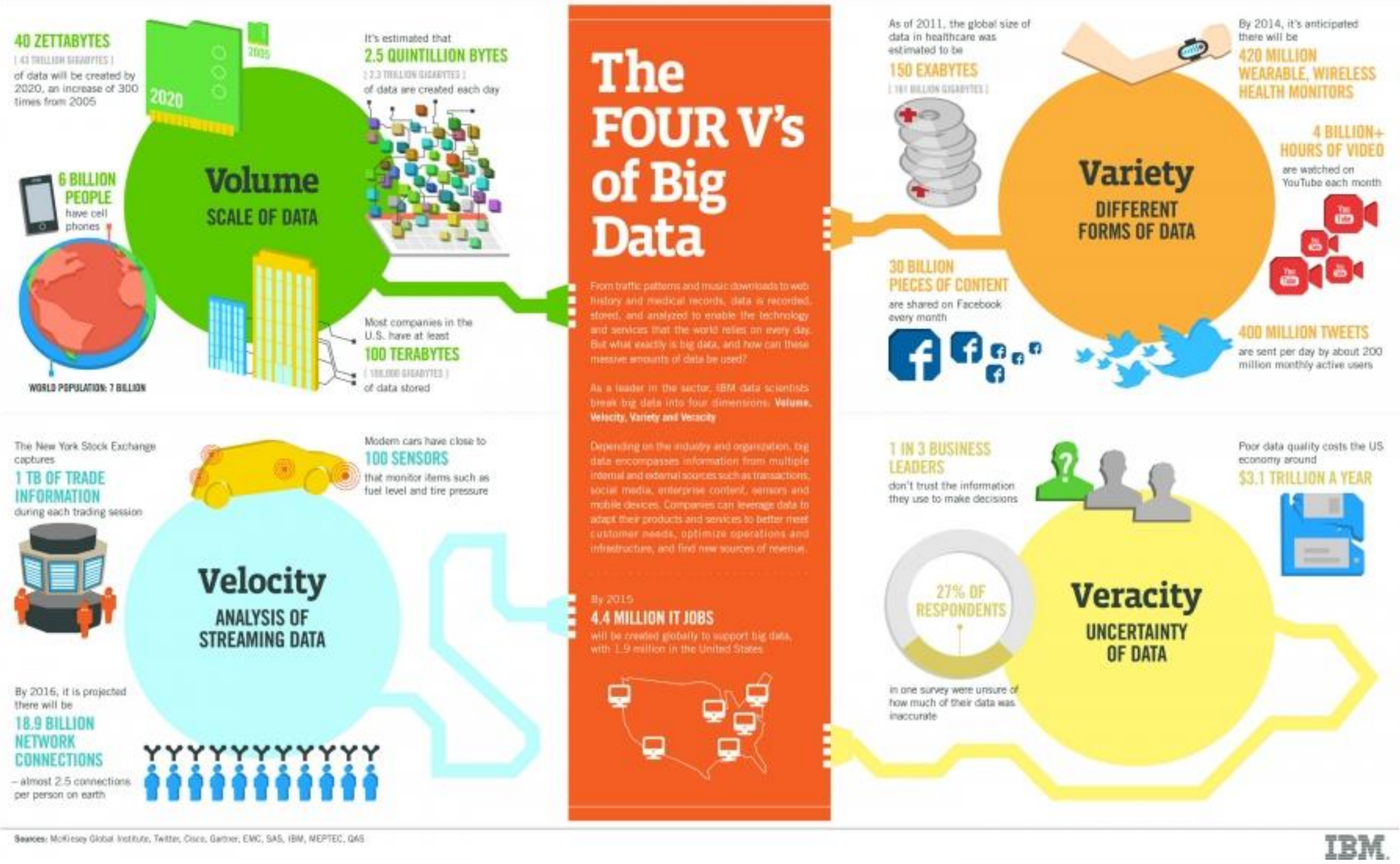
The Farr Institute of
Health Informatics Research

Big data

like teenage sex ‘everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is doing it,
so everyone claims they are doing it’

Dan Ariely

What is big data?



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States

Volume SCALE OF DATA

40 ZETTABYTES [43 TRILLION GIGABYTES] of data will be created by 2020, an increase of 300 times from 2005

It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day

6 BILLION PEOPLE have cell phones

WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

Velocity ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session

Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** – almost 2.5 connections per person on earth

Variety DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be **150 EXABYTES** [161 BILLION GIGABYTES]

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

30 BILLION PIECES OF CONTENT are shared on Facebook every month

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month

400 MILLION TWEETS are sent per day by about 200 million monthly active users

Veracity UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions

Poor data quality costs the US economy around **$3.1 TRILLION A YEAR**

27% OF RESPONDENTS in one survey were unsure of how much of their data was inaccurate

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

How do we scale science in record linkages?

National sources of health record data

NCRS eg
- Personal Demographics Service
- Personal Spine Information Service
- Transaction Messaging Service
- Secondary Uses System
- NN48 / Central Issuing System
- Choose & Book, Payment by Results, GP2GP, etc
- Electronic Prescriptions

NHS National Collections eg
- Commissioning Datasets
- Mental Health Minimum Data
- QOF / QMAS

Specialist Collections eg
Cancer / Diabetes/ Renal
Audit; waiting times; workforce

Office of National Statistics
- Births, Deaths, Terminations, Marriages
- Census & Special Surveys
- Eg HSE, NDNS, GHS CEMACH, CEPOD, Infant Feeding, etc

UNIQUE IDENTIFIER
@ BIRTH / ARRIVAL in UK
NHS NUMBER
CHILD MOTHER
FATHER

Primary care
GPRD, EMIS, THIN et al.
Child health records
Immunisations [COVER]

Hospital Care Records
Hospital admissions
Operative procedures
A&E/day cases
HES maternity 'tail
Special clinics & services
Fertility (NHS/private)
Genitourinary medicine
Intensive care
Networks

Devices/prescribing
Cochlear implants
Hip/knee replacement
MHRA systems

Diagnostic/Imaging
Ultrasound/Xray [PACS]
Mammography
Cytology/Pathology
Haematology
Chemical Pathology
Virology/Microbiology
Blood Transfusion
Screening programmes
HPA surveillance

Registers/databases
Cancer registers
Diabetes registers
Renal registers
Congenital anomaly registers
HFEA database
Cerebral palsy registers
Down syndrome register
Congenital rubella register
HIV database
Newborn screening databases
NICOR
Juvenile chronic arthritis
Inflammatory bowel disease
Dysmorphology database
Rare disorders
Public Health Observatories

Cohorts/Biobanks
1946 1958 1970 Millennium
ALSPAC, ELSA
MidSpan, Aberdeen, Walker
Generation Scotland
UK Biobank
Newborn Biobank

Environment
UK Air Quality Archive
Environmental Agency [Landfill]
Drinking Water Inspectorate
British Geological Survey
GIS data [mobile phone masts]
Superoutput areas/small area microdata

Social Care
Child Protection, In Care/ Adopted
Elderly Care

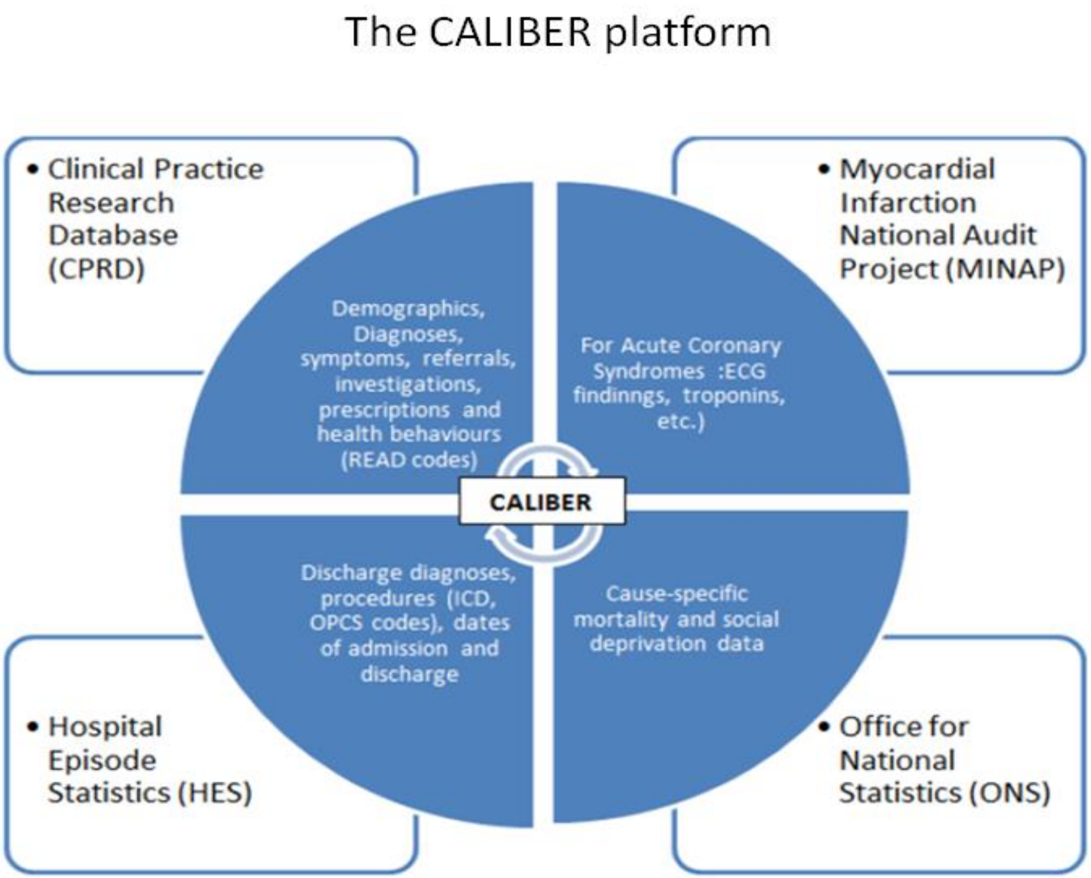
Income & Benefits
Benefits, Housing, Income

Education & Employment
Preschool/day care
Special Educational Needs
Pupil Level Annual School Census [PLASC] eg SATS scores
GCSE, GCE, Higher education
Occupations and Employment

Not available in US or Scandinavia

But many UK national record sources are not linked

Multiple Record Linkages…needs expansion across NICOR registries



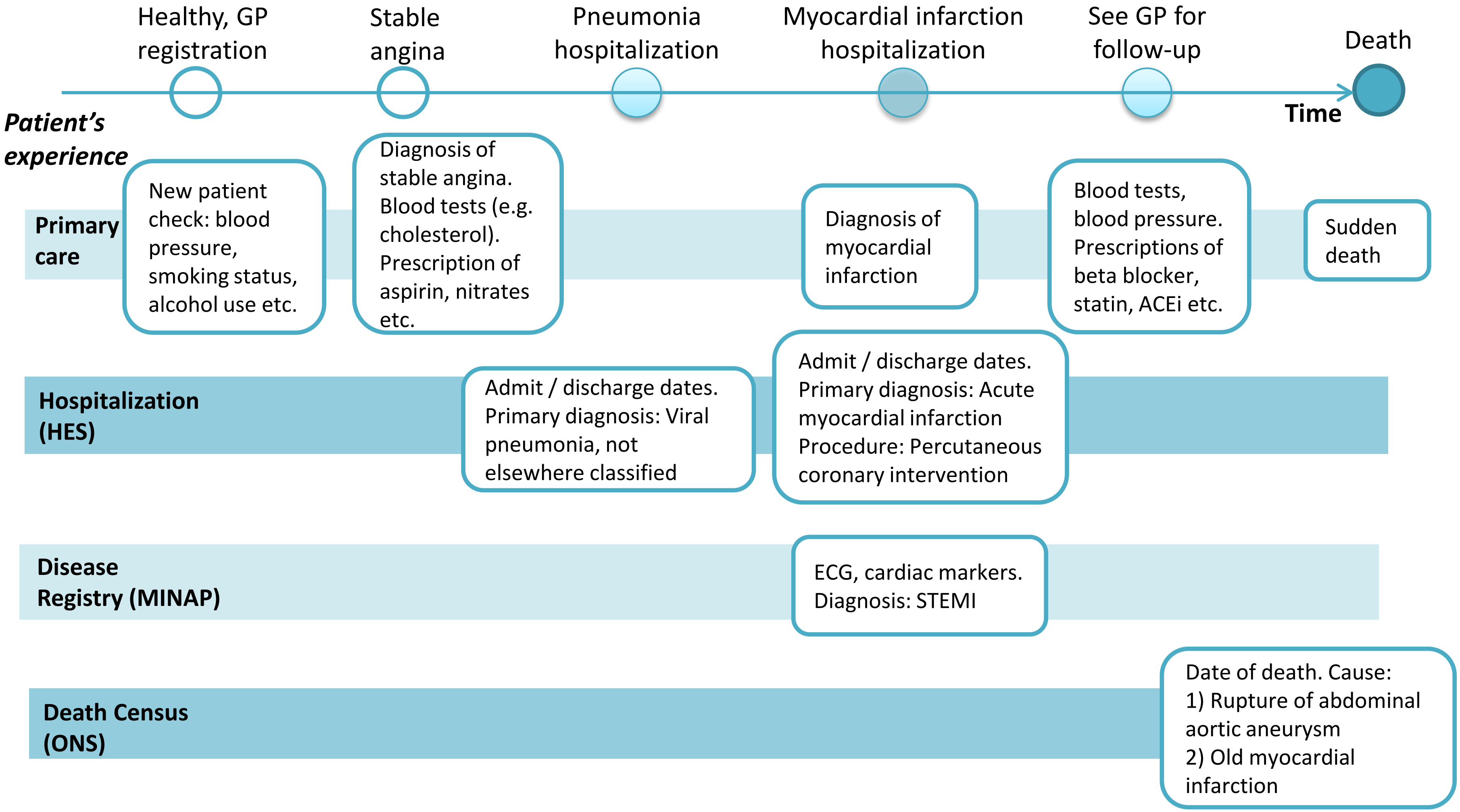
The CALIBER platform

- Clinical Practice Research Database (CPRD)
- Myocardial Infarction National Audit Project (MINAP)
- Hospital Episode Statistics (HES)
- Office for National Statistics (ONS)

Demographics, Diagnoses, symptoms, referrals, investigations, prescriptions and health behaviours (READ codes)
For Acute Coronary Syndromes :ECG findinngs, troponins, etc.)
CALIBER
Discharge diagnoses, procedures (ICD, OPCS codes), dates of admission and discharge
Cause-specific mortality and social deprivation data

The Farr Institute of Health Informatics Research

Four nationwide EHR sources linked



Healthy, GP registration

Stable angina

Pneumonia hospitalization

Myocardial infarction hospitalization

See GP for follow-up

Death

Time

Patient's experience

Primary care

- New patient check: blood pressure, smoking status, alcohol use etc.
- Diagnosis of stable angina. Blood tests (e.g. cholesterol). Prescription of aspirin, nitrates etc.
- Diagnosis of myocardial infarction
- Blood tests, blood pressure. Prescriptions of beta blocker, statin, ACEi etc.
- Sudden death

Hospitalization (HES)

- Admit / discharge dates. Primary diagnosis: Viral pneumonia, not elsewhere classified
- Admit / discharge dates. Primary diagnosis: Acute myocardial infarction Procedure: Percutaneous coronary intervention

Disease Registry (MINAP)

- ECG, cardiac markers. Diagnosis: STEMI

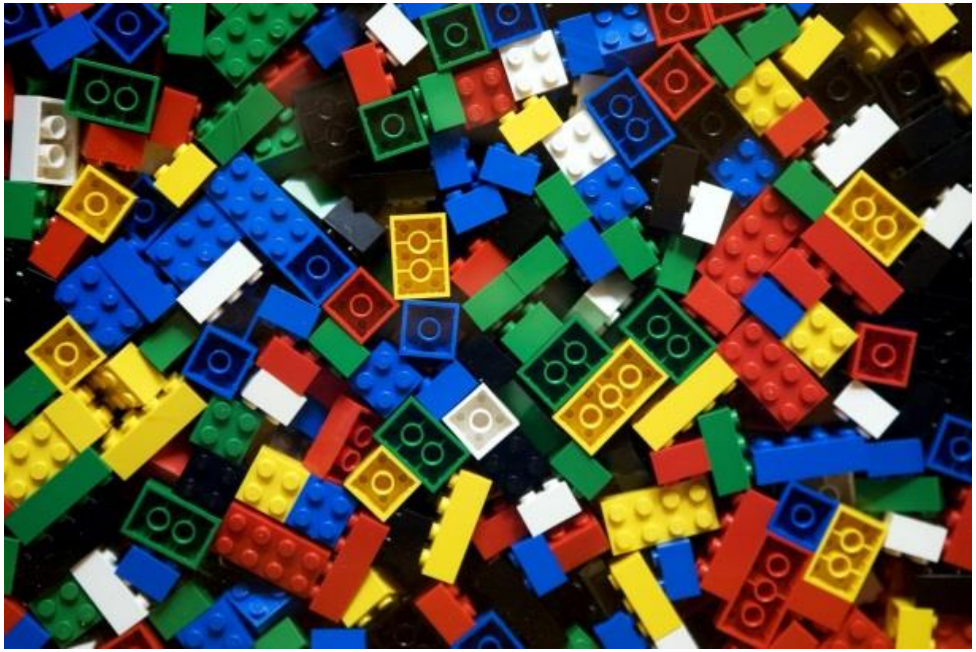
Death Census (ONS)

- Date of death. Cause:
  1) Rupture of abdominal aortic aneurysm
  2) Old myocardial infarction

Denaxas et al, CALIBER, Intl J Epidemiology, 2012;41(6):1625-38.

The Farr Institute of Health Informatics Research

What does linked record data look like?

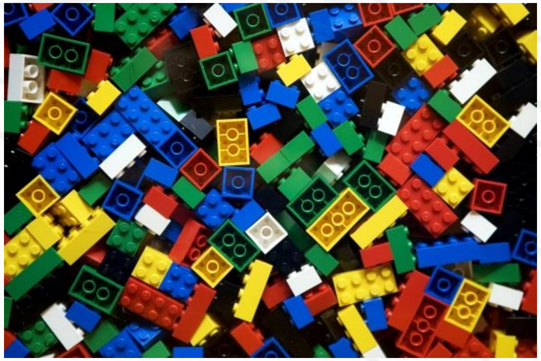


To get at big data…need tools

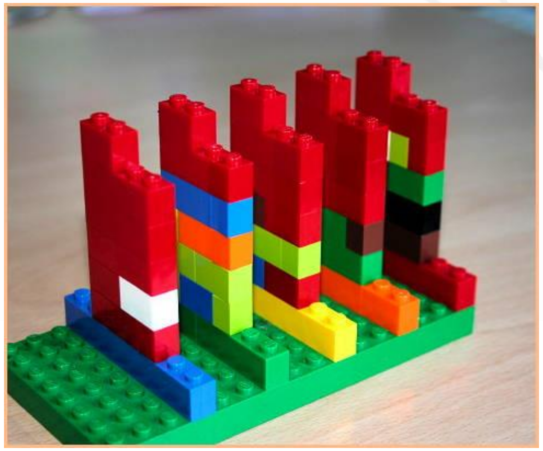
How to define phenotypes using multiple EHR data sources?

ATRIAL FIBRILLATION AND FLUTTER

```
1001, 2000-01-01, 23,1,NULL,I48
1001, 1994-08-11,1234,1,3,7L1H300
1001, 1993-01-01, 253,1,1,793Mz00
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, 1991-03-03,1234,1,3,7L1H310
682444, 1993-01-01, 253,1,1,793Mz00
```

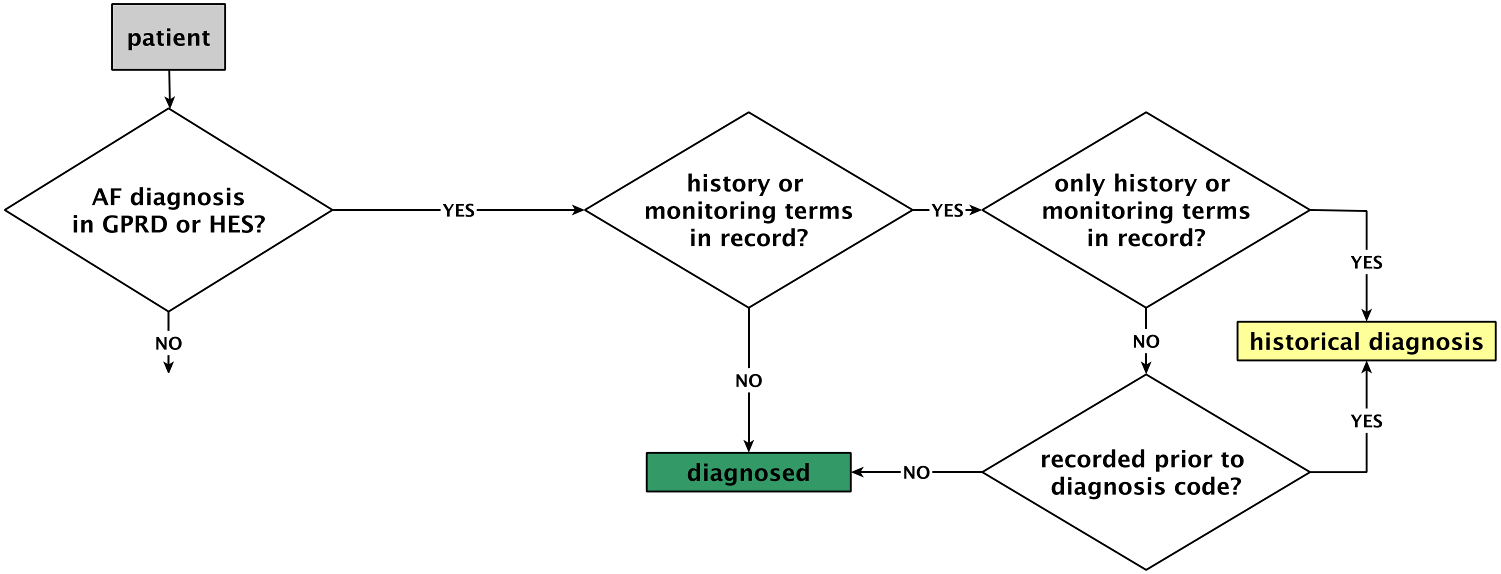


1001, 2000-01-01, **af_gprd=1**
1231, 2012-03-03, **af_hes=3**
1121, 2013-05-04, **af_procs_gprd=1**
1511, 1993-01-11, **heart_valve_gprd=2**
9913, 2012-05-21, **af_hes=1**
67222, 1994-08-11, **af_hes=1**
682444, 1993-01-01, **heart_valve_hes=2**



af=1, af_diag_source="primary care" af_diag_date=2001-12-01

AF algorithm



patient

AF diagnosis in GPRD or HES?

YES

NO

history or monitoring terms in record?

YES

NO

only history or monitoring terms in record?

YES

NO

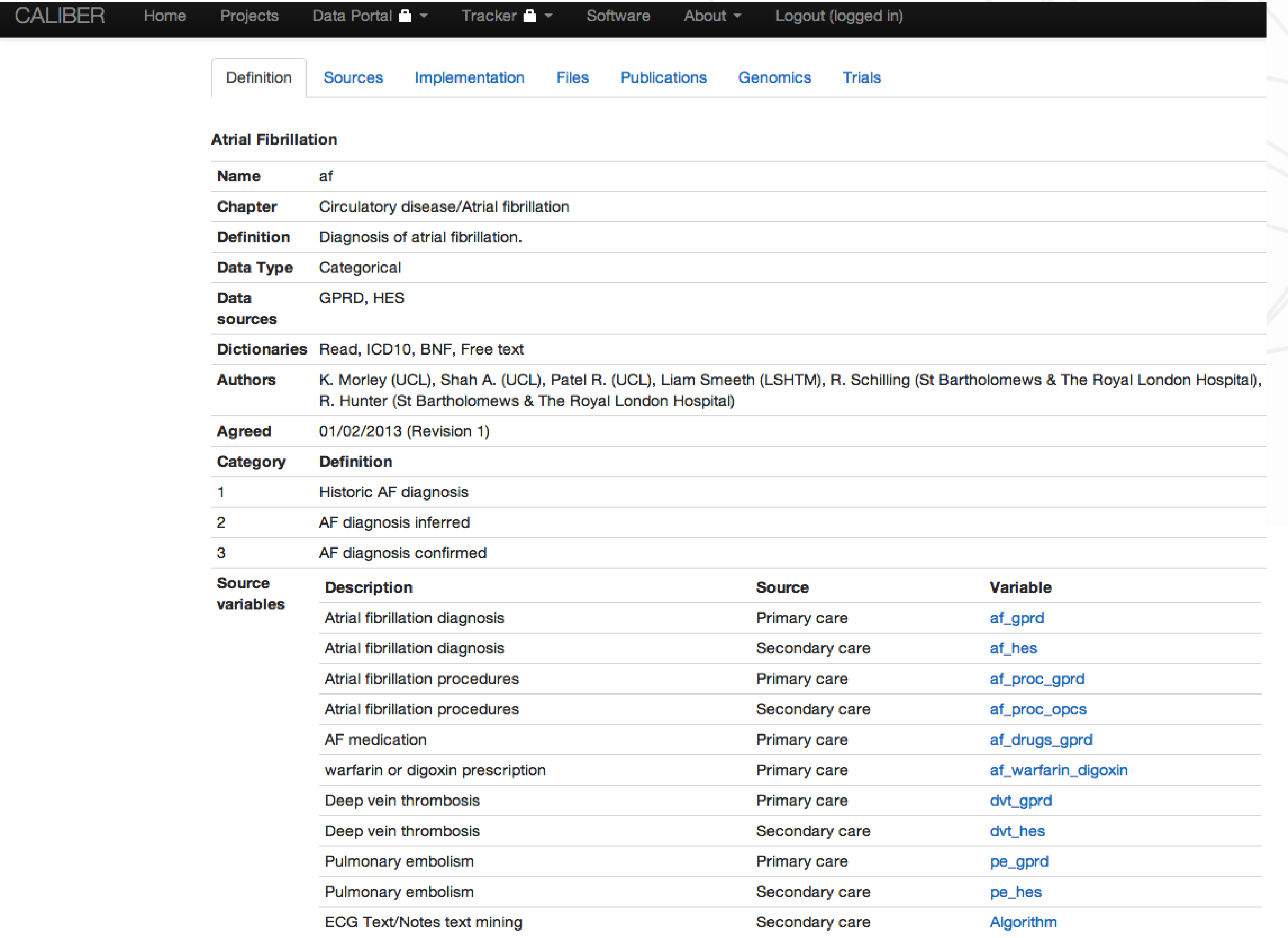
historical diagnosis

diagnosed

recorded prior to diagnosis code?

NO

YES



Atrial Fibrillation

Name	af
Chapter	Circulatory disease/Atrial fibrillation
Definition	Diagnosis of atrial fibrillation.
Data Type	Categorical
Data sources	GPRD, HES
Dictionaries	Read, ICD10, BNF, Free text
Authors	K. Morley (UCL), Shah A. (UCL), Patel R. (UCL), Liam Smeeth (LSHTM), R. Schilling (St Bartholomews & The Royal London Hospital), R. Hunter (St Bartholomews & The Royal London Hospital)
Agreed	01/02/2013 (Revision 1)

Category	**Definition**
1	Historic AF diagnosis
2	AF diagnosis inferred
3	AF diagnosis confirmed

Source variables

Description	**Source**	**Variable**
Atrial fibrillation diagnosis	Primary care	af_gprd
Atrial fibrillation diagnosis	Secondary care	af_hes
Atrial fibrillation procedures	Primary care	af_proc_gprd
Atrial fibrillation procedures	Secondary care	af_proc_opcs
AF medication	Primary care	af_drugs_gprd
warfarin or digoxin prescription	Primary care	af_warfarin_digoxin
Deep vein thrombosis	Primary care	dvt_gprd
Deep vein thrombosis	Secondary care	dvt_hes
Pulmonary embolism	Primary care	pe_gprd
Pulmonary embolism	Secondary care	pe_hes
ECG Text/Notes text mining	Secondary care	Algorithm

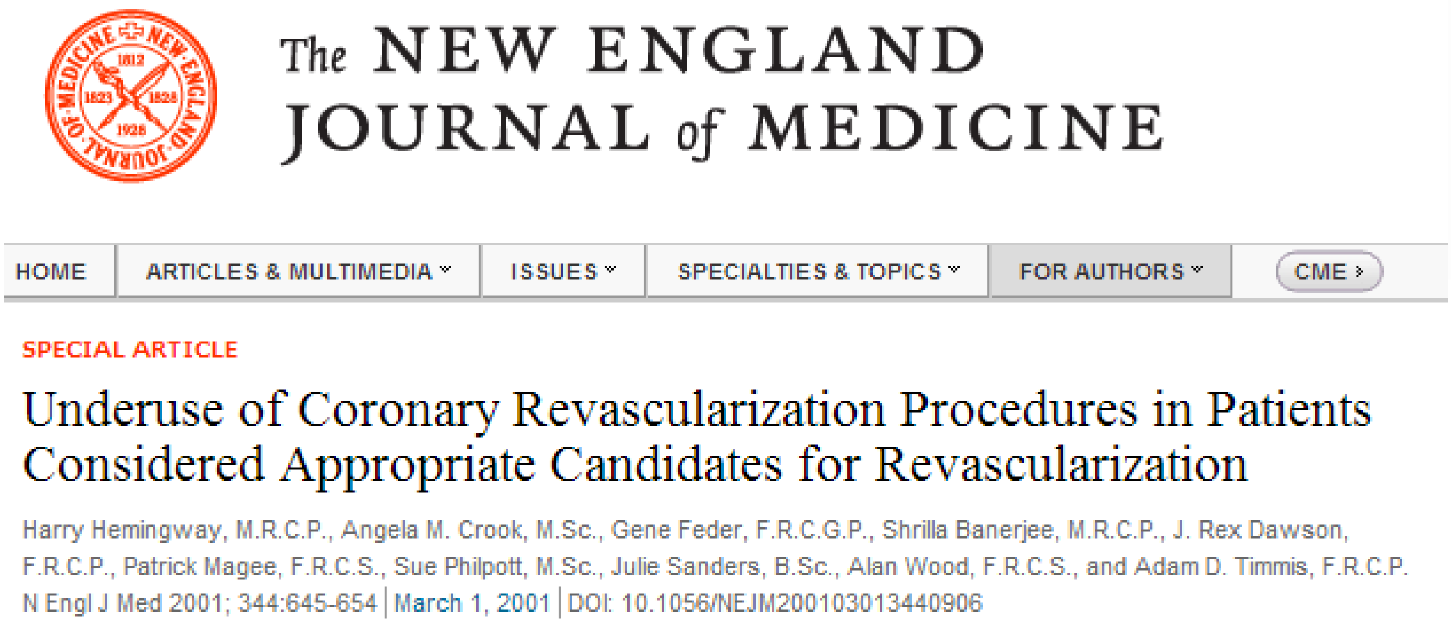
CALIBER Data Portal

- Online data discovery tool **caliberresearch.org**
- Access to *all* CALIBER phenotypes, algorithms and implementation details and scripts (SQL,R, Stata)
  - 45 users, 4 institutions, 538 phenotypes, >15,000 clinical diagnostic codes curated
- Standardization
  - Frontend is ICD10, backend becoming SNOMED-CT, LOINC
- A community rather than a static resource
  - Researchers contribute phenotypes and algorithms
  - Other researchers validate/enhance/correct them

The Farr Institute of Health Informatics Research

Why me?

Why me?



The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

Underuse of Coronary Revascularization Procedures in Patients Considered Appropriate Candidates for Revascularization

Harry Hemingway, M.R.C.P., Angela M. Crook, M.Sc., Gene Feder, F.R.C.G.P., Shrilla Banerjee, M.R.C.P., J. Rex Dawson, F.R.C.P., Patrick Magee, F.R.C.S., Sue Philpott, M.Sc., Julie Sanders, B.Sc., Alan Wood, F.R.C.S., and Adam D. Timmis, F.R.C.P.

N Engl J Med 2001; 344:645-654 | March 1, 2001 | DOI: 10.1056/NEJM200103013440906

The Farr Institute of
Health Informatics Research

What's wrong that big data might help fix?

The Farr Institute of
Health Informatics Research

Cardiovascular diseases global #1

- Cause of death/premature death/disability adjusted life years

- So what has gone wrong?

  - wrong prevention
  - wrong treatments
  - wrong diagnoses / wrong names for diseases
  - wrong health systems (and too costly)
  - wrong relations to data, information and knowledge
  - wrong relations with patients

  - wrong science! (done by the wrong people!)

Might big data help right these wrongs?

The Farr Institute of
Health Informatics Research

Yes!

Mike Lauer, Director Division of Cardiovascular Sciences National Institutes of Health

EDITORIAL

Editorials represent the opinions of the authors and JAMA and not those of the American Medical Association.

Time for a Creative Transformation of Epidemiology in the United States

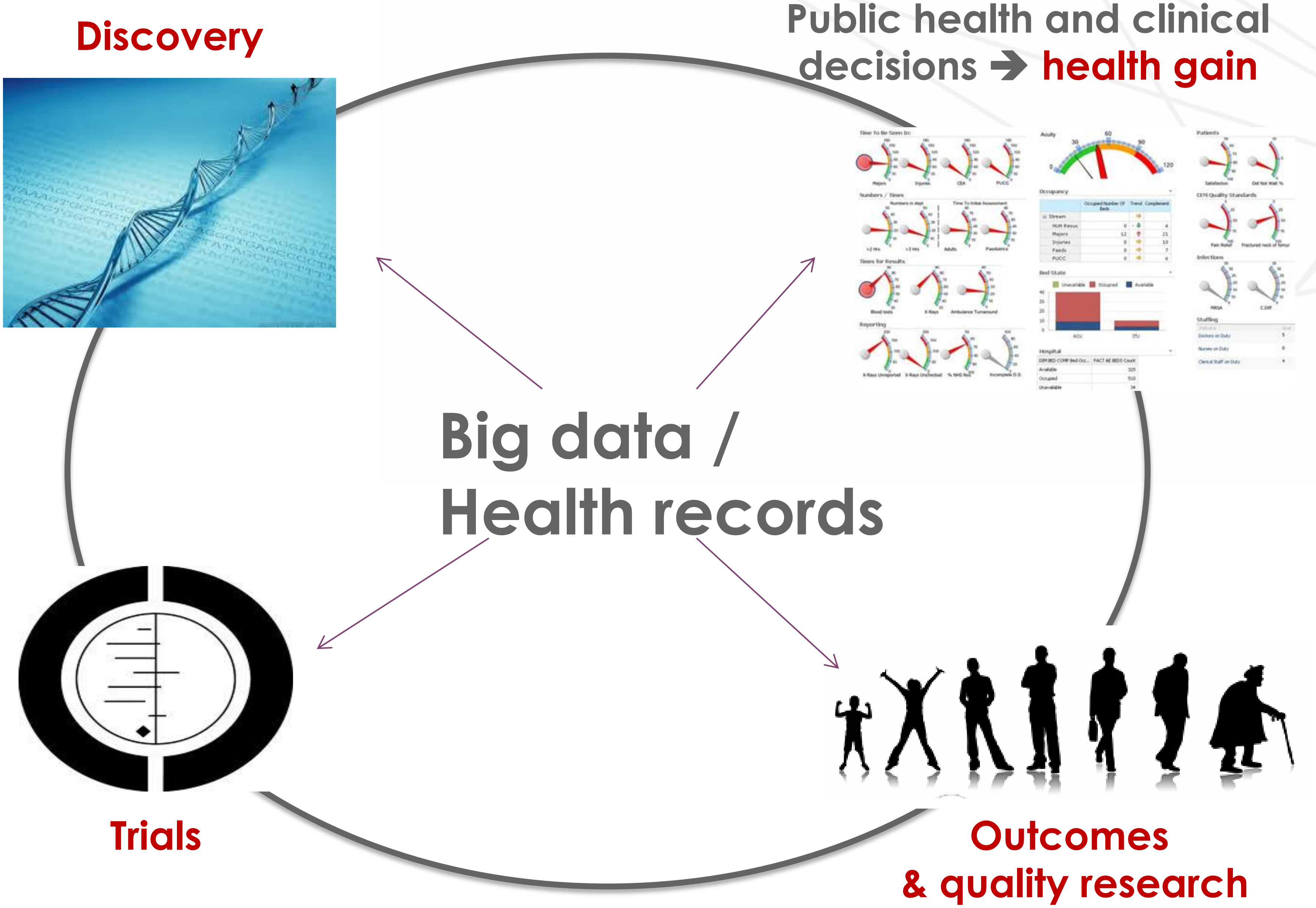
JAMA 2012

"US models are being eclipsed by non-US studies that are much larger, yet considerably less expensive"

Yes!

Eric Topol ‘wireless and genomic medicine’

Pace and scale of translation



Discovery

Public health and clinical decisions → **health gain**

Big data /
Health records

Trials

**Outcomes
& quality research**

Discovery

The Farr Institute of
Health Informatics Research

Genomics

biobank uk

500k participants, 47 baseline biomarkers and custom gene array data available in 2014, cardiac and brain imaging in 100k underway

Open access

Scalable approaches to disease phenotypes (startpoints or endpoints) based on linked electronic health record resources

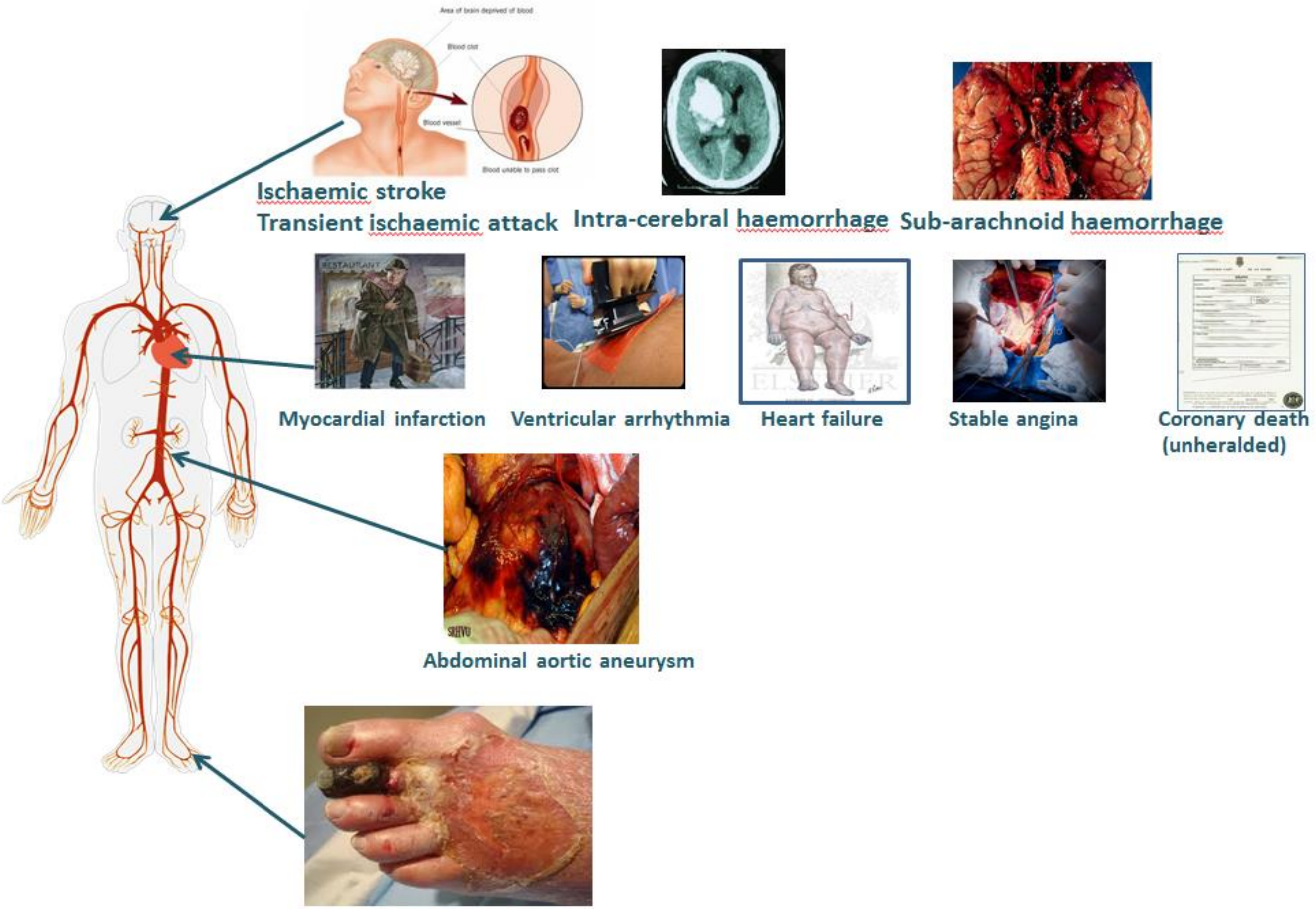
- cardiac
- diabetes
- stroke
- cancer

Example of Farr Institute working across Wales, Scotland and England

Discovering new risk factor associations:

CVD aggregates vs specific diseases

Are the risk factors the same?



Ischaemic stroke
Transient ischaemic attack

Intra-cerebral haemorrhage

Sub-arachnoid haemorrhage

Myocardial infarction

Ventricular arrhythmia

Heart failure

Stable angina

Coronary death (unheralded)

Abdominal aortic aneurysm

Peripheral arterial disease

The Farr Institute of Health Informatics Research

To answer this question reliably we need

- **Scale:** e.g. >1 million adults followed for 5 years

- **Phenotypic resolution:**
  - Baseline risk factors
  - Follow up for disease outcomes

Cost to research funder of such data collection?

The Farr Institute of
Health Informatics Research

£0.00

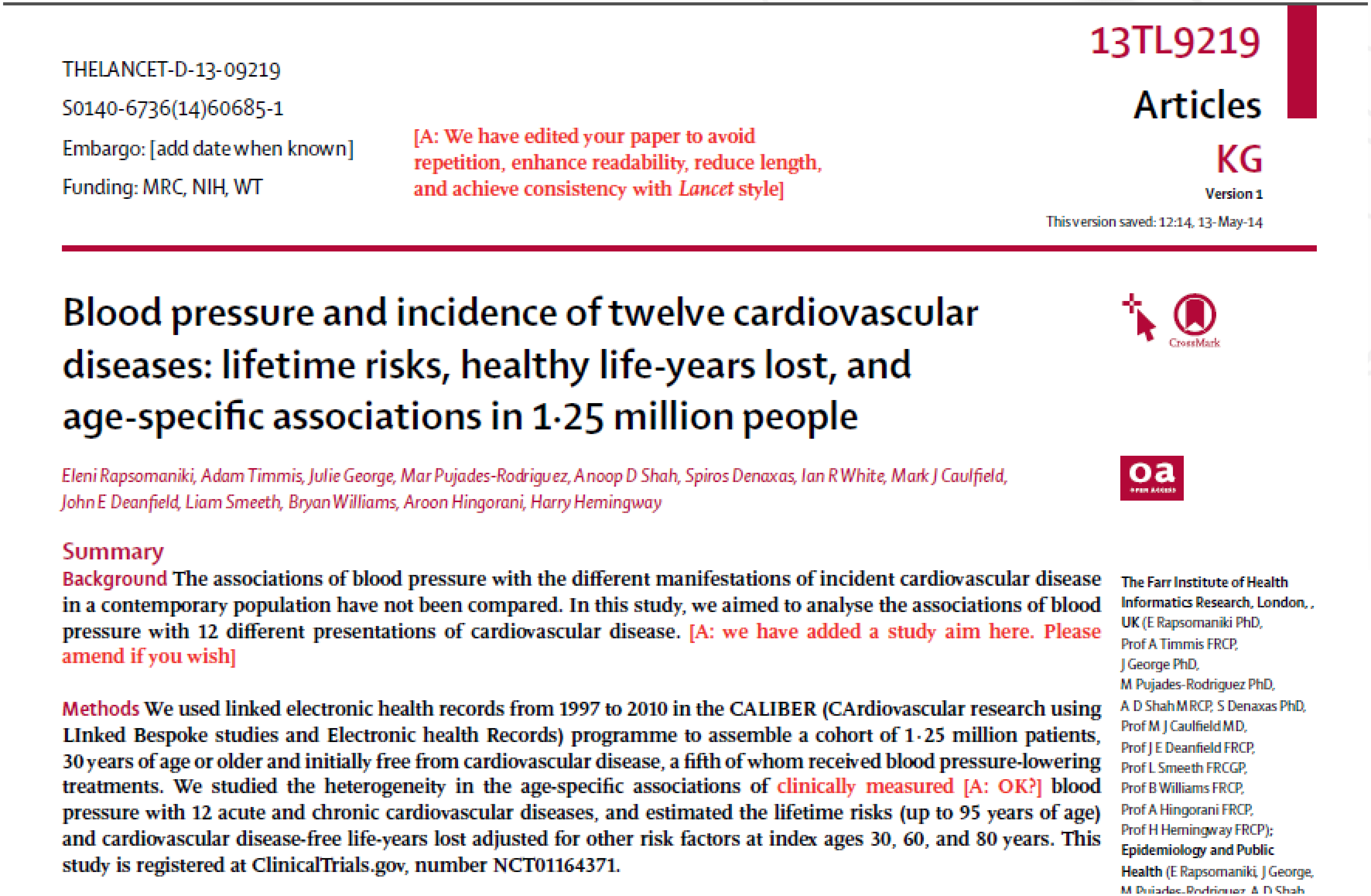
The research costs <u>are substantial</u>

Information governance

Store, share, harmonise, analyse EHR data.....with scalable tools

<u>And develop pool of clinical expertise</u>

The Farr Institute of
Health Informatics Research



THELANCET-D-13-09219

S0140-6736(14)60685-1

Embargo: [add date when known]

Funding: MRC, NIH, WT

[A: We have edited your paper to avoid repetition, enhance readability, reduce length, and achieve consistency with *Lancet* style]

13TL9219

Articles

KG

Version 1

This version saved: 12:14, 13-May-14

Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people

Eleni Rapsomaniki, Adam Timmis, Julie George, Mar Pujades-Rodriguez, Anoop D Shah, Spiros Denaxas, Ian R White, Mark J Caulfield, John E Deanfield, Liam Smeeth, Bryan Williams, Aroon Hingorani, Harry Hemingway

oa

Summary

Background The associations of blood pressure with the different manifestations of incident cardiovascular disease in a contemporary population have not been compared. In this study, we aimed to analyse the associations of blood pressure with 12 different presentations of cardiovascular disease. [A: we have added a study aim here. Please amend if you wish]

Methods We used linked electronic health records from 1997 to 2010 in the CALIBER (CArdiovascular research using LInked Bespoke studies and Electronic health Records) programme to assemble a cohort of 1·25 million patients, 30 years of age or older and initially free from cardiovascular disease, a fifth of whom received blood pressure-lowering treatments. We studied the heterogeneity in the age-specific associations of clinically measured [A: OK?] blood pressure with 12 acute and chronic cardiovascular diseases, and estimated the lifetime risks (up to 95 years of age) and cardiovascular disease-free life-years lost adjusted for other risk factors at index ages 30, 60, and 80 years. This study is registered at ClinicalTrials.gov, number NCT01164371.

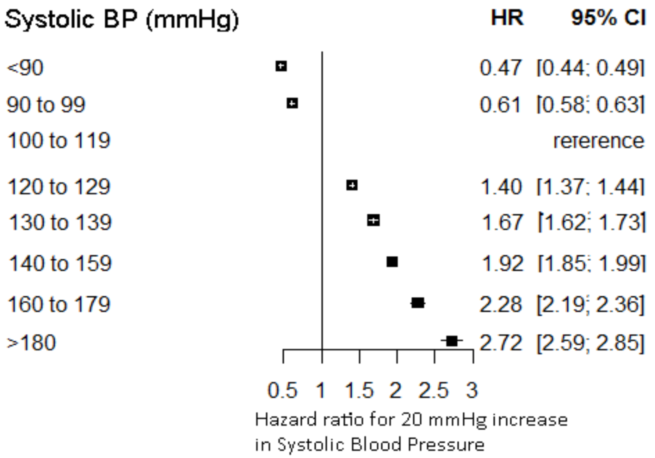
The Farr Institute of Health Informatics Research, London, , UK (E Rapsomaniki PhD, Prof A Timmis FRCP, J George PhD, M Pujades-Rodriguez PhD, A D Shah MRCP, S Denaxas PhD, Prof M J Caulfield MD, Prof J E Deanfield FRCP, Prof L Smeeth FRCGP, Prof B Williams FRCP, Prof A Hingorani FRCP, Prof H Hemingway FRCP); **Epidemiology and Public Health** (E Rapsomaniki, J George, M Pujades-Rodriguez, A D Shah

Farr
The Farr Institute of
Health Informatics Research

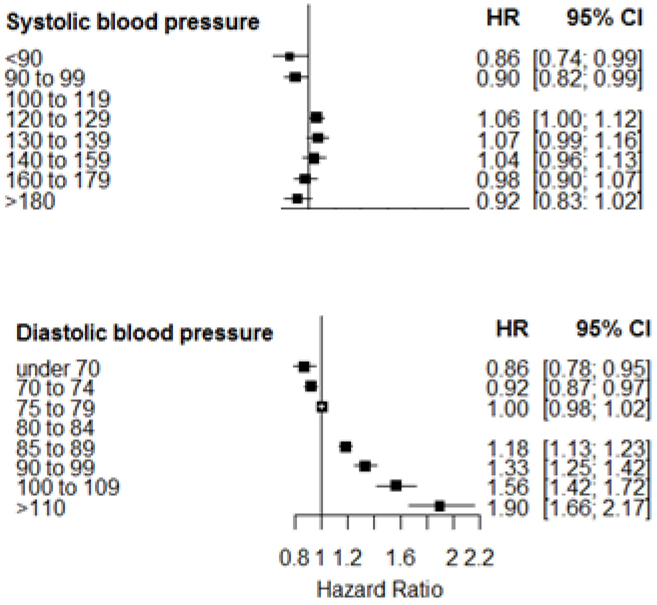
Higher resolution epidemiology: blood pressure and 12 cardiovascular diseases

Cohort N ≈ 2 million adults, >100,000 disease events

Myocardial infarction



Abdominal aortic aneurysm



Confirms what we know from combining multiple expensive studies

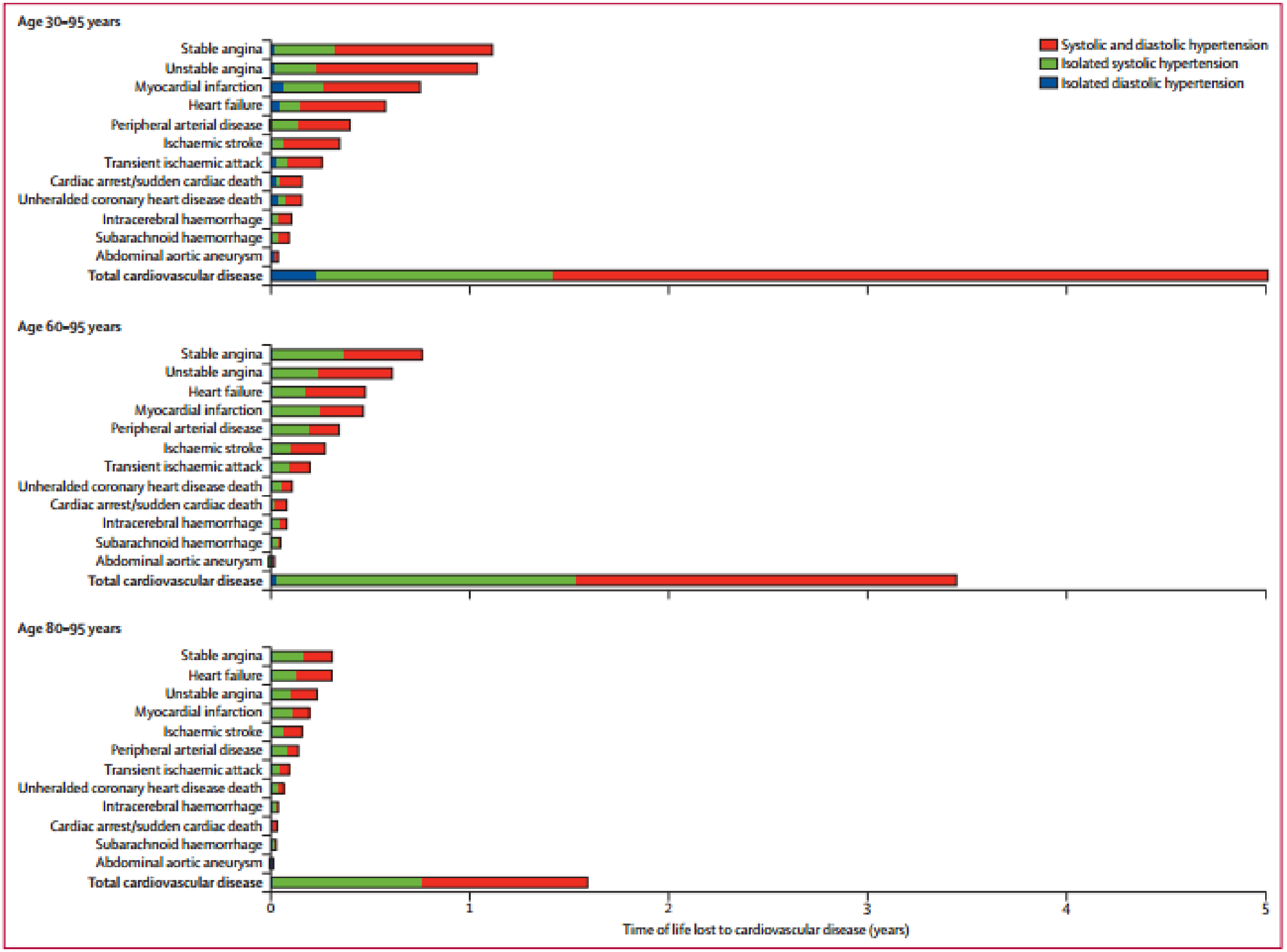
Adds resolution

New knowledge

....a challenge for experimental medicine

Rapsomaniki et al, CALIBER The Lancet 2014

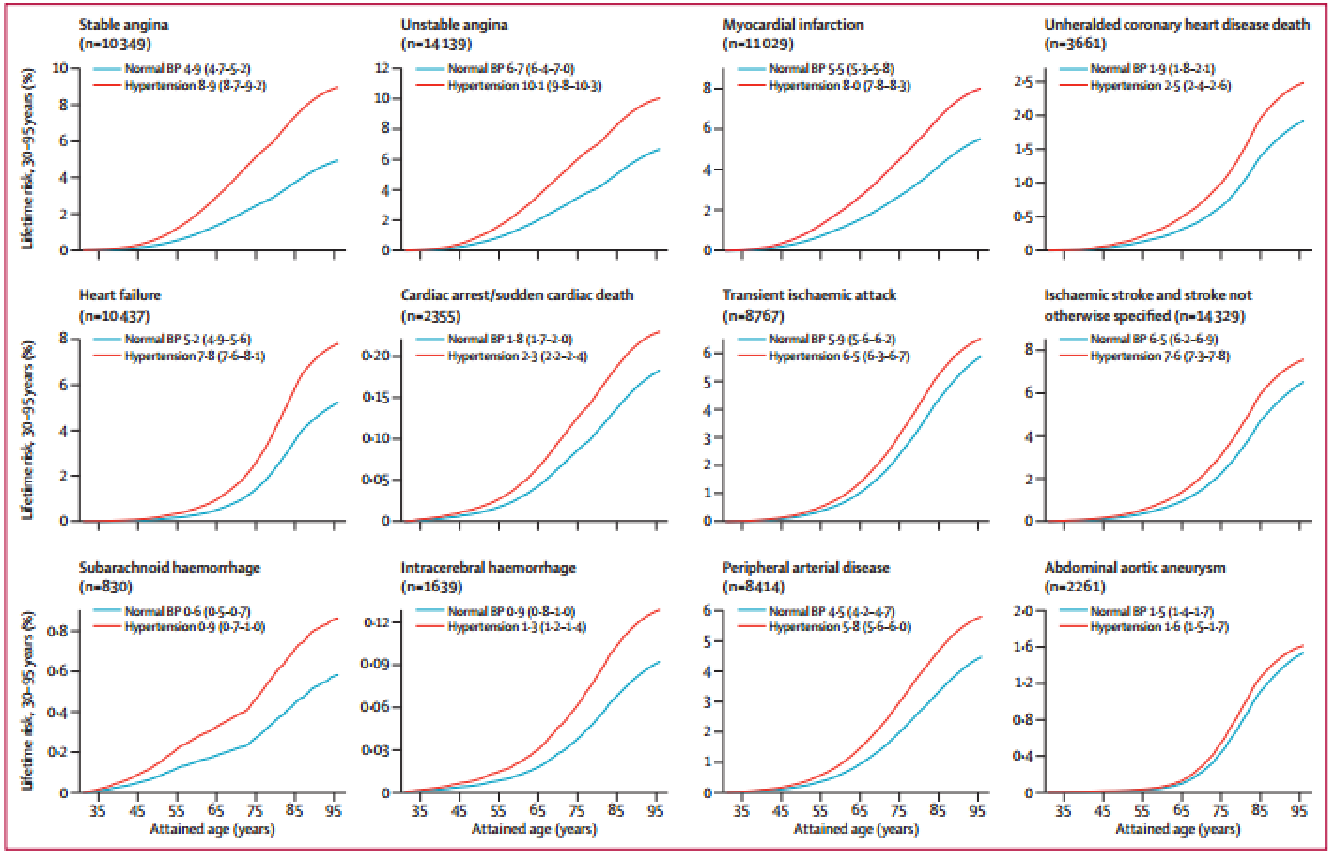
Years of life lost to CVD



Rapsomaniki et al, CALIBER Lancet 2014;383(9932):1899-911

The Farr Institute of Health Informatics Research

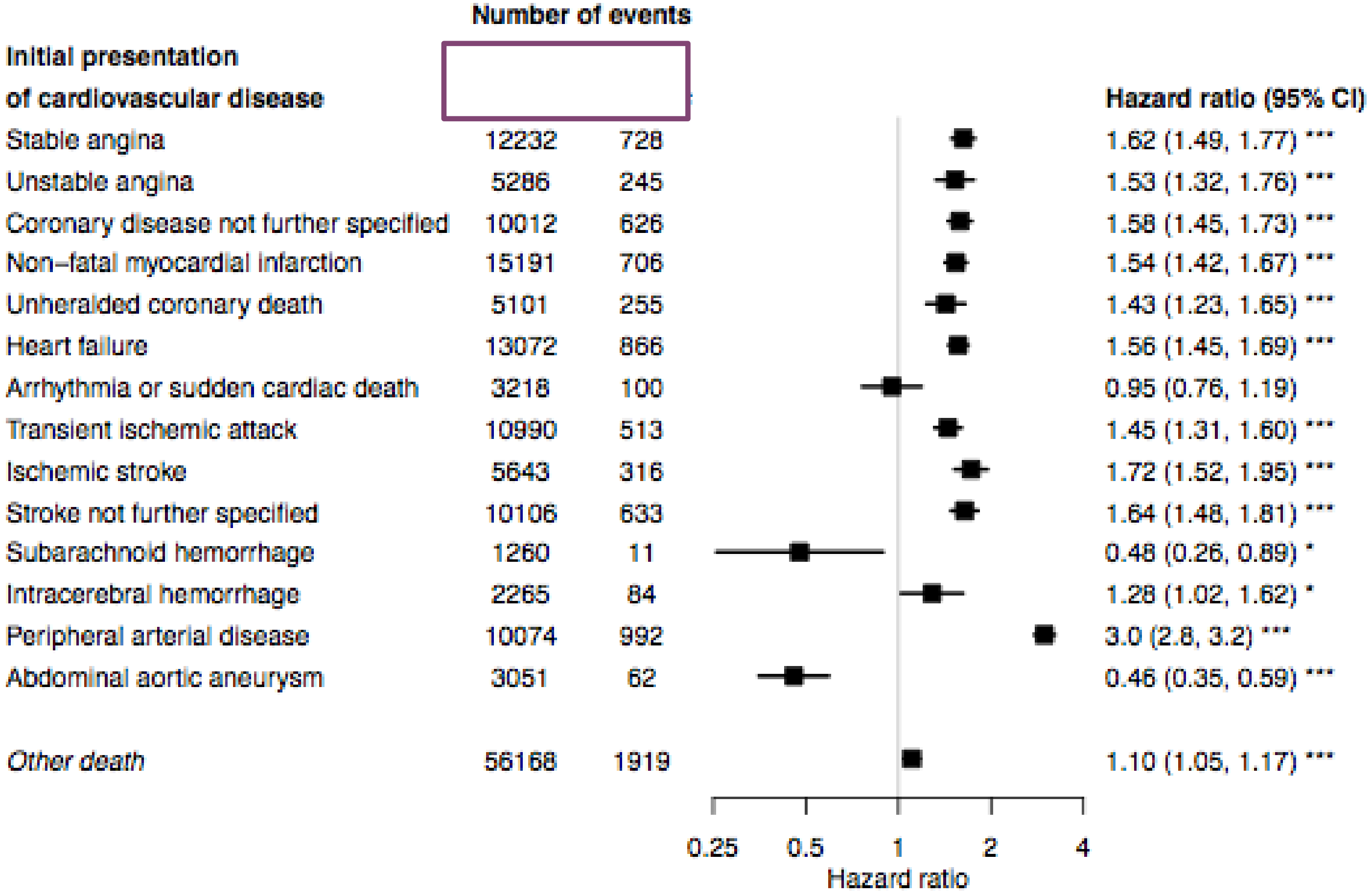
Cumulative life time risk of 12 cardiovascular diseases



Rapsomaniki et al, CALIBER Lancet 2014;383(9932):1899-911

The Farr Institute of Health Informatics Research

Inverse, null, weak and strong…what's the 'risk factor'?



Initial presentation of cardiovascular disease	Number of events		Hazard ratio (95% CI)
Stable angina	12232	728	1.62 (1.49, 1.77) ***
Unstable angina	5286	245	1.53 (1.32, 1.76) ***
Coronary disease not further specified	10012	626	1.58 (1.45, 1.73) ***
Non–fatal myocardial infarction	15191	706	1.54 (1.42, 1.67) ***
Unheralded coronary death	5101	255	1.43 (1.23, 1.65) ***
Heart failure	13072	866	1.56 (1.45, 1.69) ***
Arrhythmia or sudden cardiac death	3218	100	0.95 (0.76, 1.19)
Transient ischemic attack	10990	513	1.45 (1.31, 1.60) ***
Ischemic stroke	5643	316	1.72 (1.52, 1.95) ***
Stroke not further specified	10106	633	1.64 (1.48, 1.81) ***
Subarachnoid hemorrhage	1260	11	0.48 (0.26, 0.89) *
Intracerebral hemorrhage	2265	84	1.28 (1.02, 1.62) *
Peripheral arterial disease	10074	992	3.0 (2.8, 3.2) ***
Abdominal aortic aneurysm	3051	62	0.46 (0.35, 0.59) ***
Other death	56168	1919	1.10 (1.05, 1.17) ***

Shah et al, CALIBER, Am J Epidemiol 2014;179(6):764-74

‘Higher resolution’ approaches: implications

- Disease mechanism
- Trial design
- Screening and risk prediction

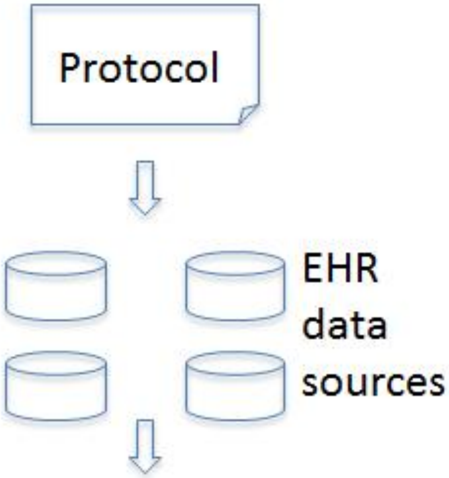
Discovery

Trials

Developing informatics platforms for stratified trials

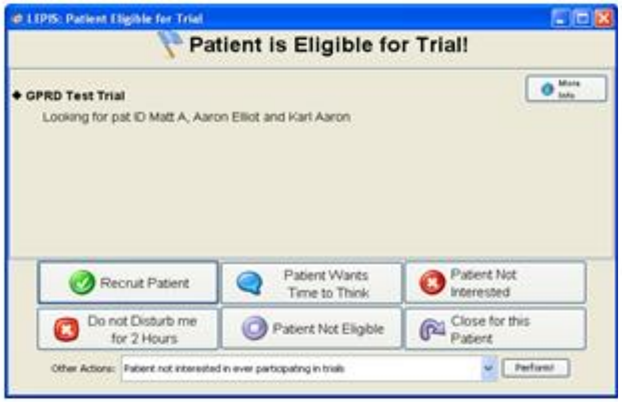
Rapid feasibility

EHR-based eligibility counts



Recruiting

EHR randomisation



LEPIS: Patient Eligible for Trial

Patient is Eligible for Trial!

◆ **GPRD Test Trial**

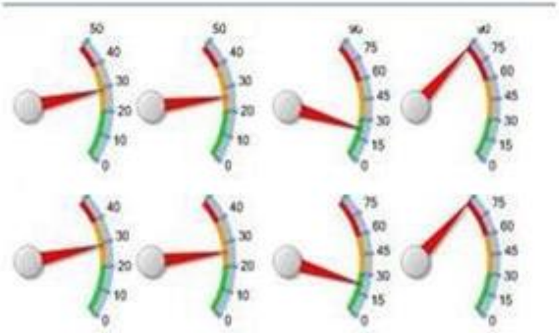
Looking for pat ID Matt A, Aaron Elliot and Karl Aaron

Other Actions: Patient not interested in ever participating in trials

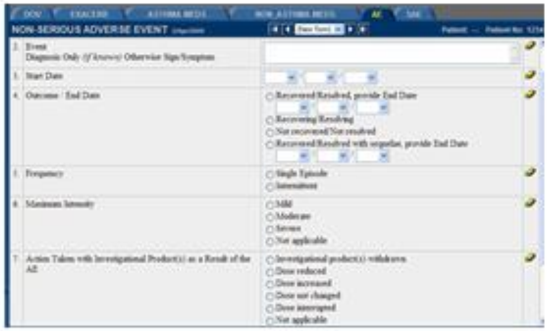
UCLP eConsent

Following up & safety

Real-time outcome dashboards

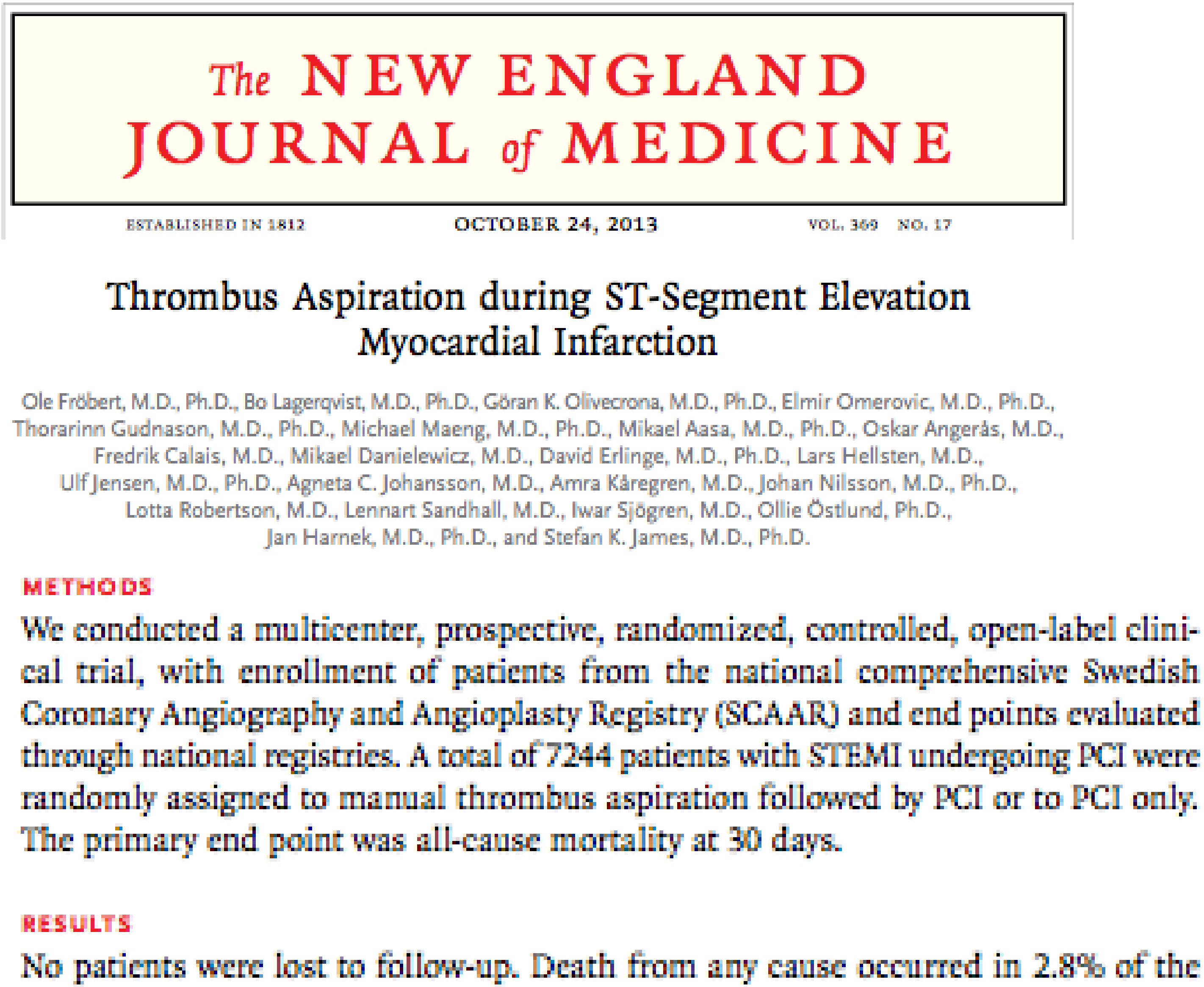


Embedded eCRF



NON-SERIOUS ADVERSE EVENT

2.	Event Diagnosis Only *(if known)* Otherwise Sign/Symptom	
3.	Start Date	
4.	Outcome / End Date	○ Recovered/Resolved, provide End Date ○ Recovering/Resolving ○ Not recovered/Not resolved ○ Recovered/Resolved with sequelae, provide End Date
5.	Frequency	○ Single Episode ○ Intermittent
6.	Maximum Intensity	○ Mild ○ Moderate ○ Severe ○ Not applicable
7.	Action Taken with Investigational Product(s) as a Result of the AE	○ Investigational product(s) withdrawn ○ Dose reduced ○ Dose increased ○ Dose not changed ○ Dose interrupted ○ Not applicable



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 OCTOBER 24, 2013 VOL. 369 NO. 17

Thrombus Aspiration during ST-Segment Elevation Myocardial Infarction

Ole Fröbert, M.D., Ph.D., Bo Lagerqvist, M.D., Ph.D., Göran K. Olivecrona, M.D., Ph.D., Elmir Omerovic, M.D., Ph.D., Thorarinn Gudnason, M.D., Ph.D., Michael Maeng, M.D., Ph.D., Mikael Aasa, M.D., Ph.D., Oskar Angerås, M.D., Fredrik Calais, M.D., Mikael Danielewicz, M.D., David Erlinge, M.D., Ph.D., Lars Hellsten, M.D., Ulf Jensen, M.D., Ph.D., Agneta C. Johansson, M.D., Amra Kåregren, M.D., Johan Nilsson, M.D., Ph.D., Lotta Robertson, M.D., Lennart Sandhall, M.D., Iwar Sjögren, M.D., Ollie Östlund, Ph.D., Jan Harnek, M.D., Ph.D., and Stefan K. James, M.D., Ph.D.

METHODS

We conducted a multicenter, prospective, randomized, controlled, open-label clinical trial, with enrollment of patients from the national comprehensive Swedish Coronary Angiography and Angioplasty Registry (SCAAR) and end points evaluated through national registries. A total of 7244 patients with STEMI undergoing PCI were randomly assigned to manual thrombus aspiration followed by PCI or to PCI only. The primary end point was all-cause mortality at 30 days.

RESULTS

No patients were lost to follow-up. Death from any cause occurred in 2.8% of the

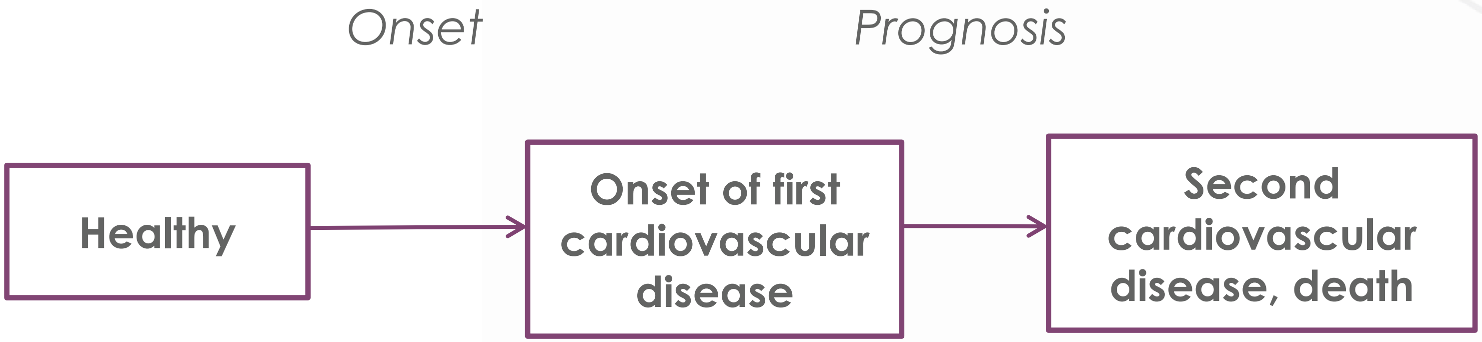
Discovery

Trials

Outcomes research/real world evidence

Temporal resolution

….with ‘big data’ can study both



Onset

Prognosis

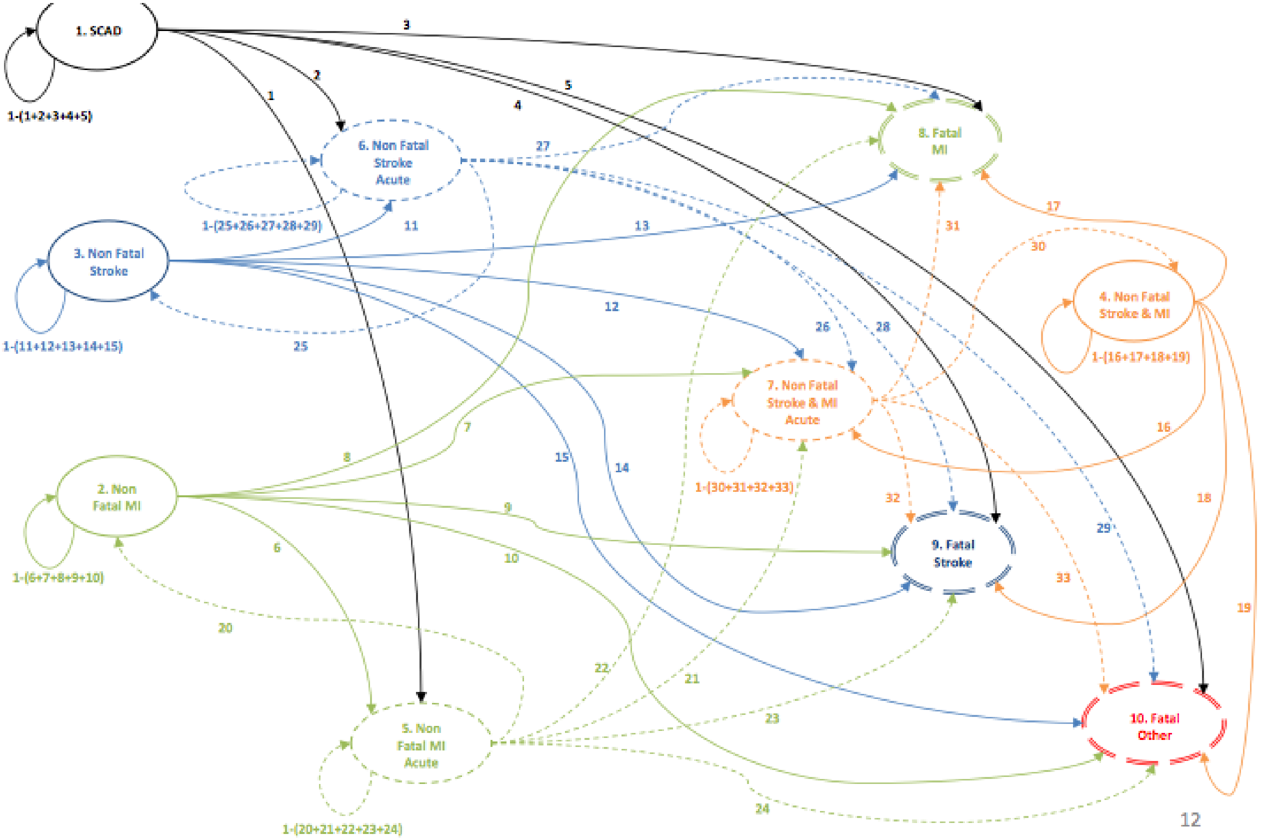
Healthy

Onset of first cardiovascular disease

Second cardiovascular disease, death

Outcomes research: capturing clinically meaningful complexity

one startpoint to many types of endpoint



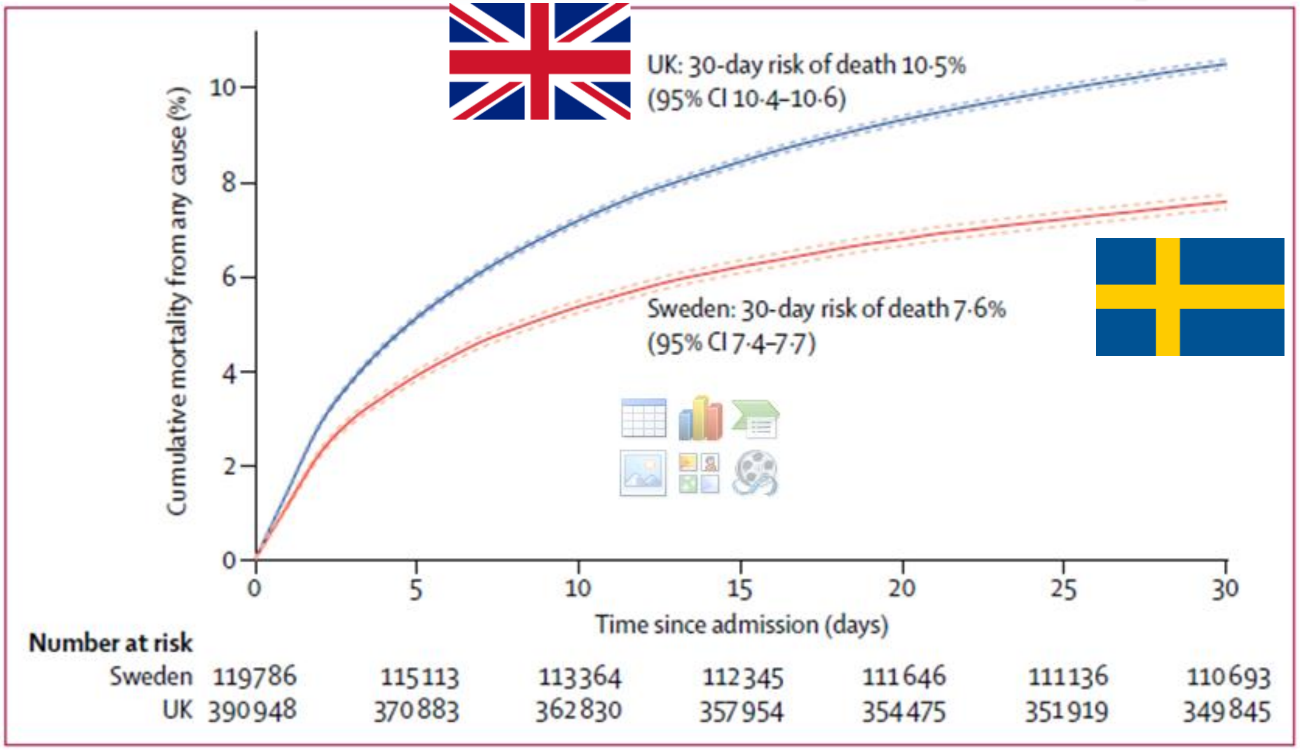
PROGRESS 4 article series in BMJ / PLoSMed 2013

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

Sheng-Chia Chung, Rolf Gedeborg, Owen Nicholas, Stefan James, Anders Jeppsson, Charles Wolfe, Peter Heuschmann, Lars Wallentin, John Deanfield, Adam Timmis, Tomas Jernberg, Harry Hemingway

Lancet 2014; 383: 1305–12

THE LANCET



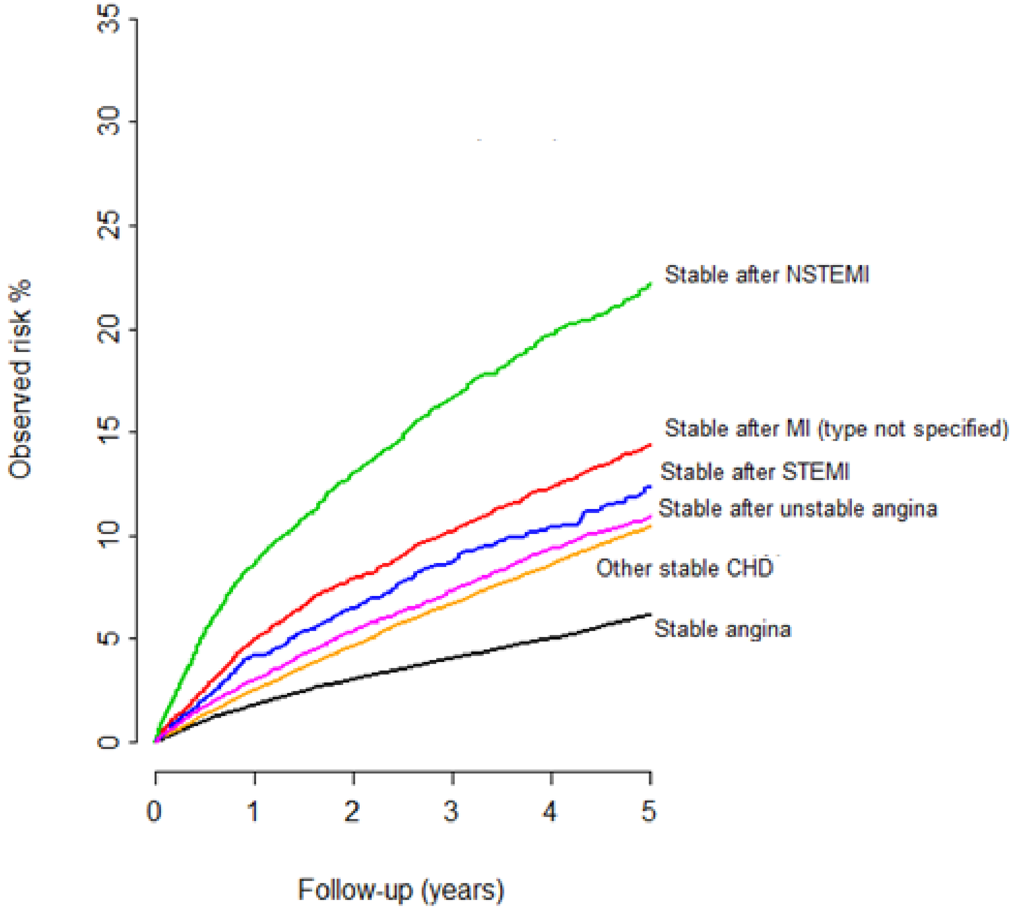
Number at risk

	0	5	10	15	20	25	30
Sweden	119786	115113	113364	112345	111646	111136	110693
UK	390948	370883	362830	357954	354475	351919	349845

Figure 3: Kaplan-Meier curves for cumulative mortality at 30 days after admission with acute myocardial infarction in Sweden and the UK

The Farr Institute of
Health Informatics Research

‘Real world’ prognosis of stable CAD
(n=102, 023) and 5 yr risk of coronary death + non-fatal MI (n=8,856)

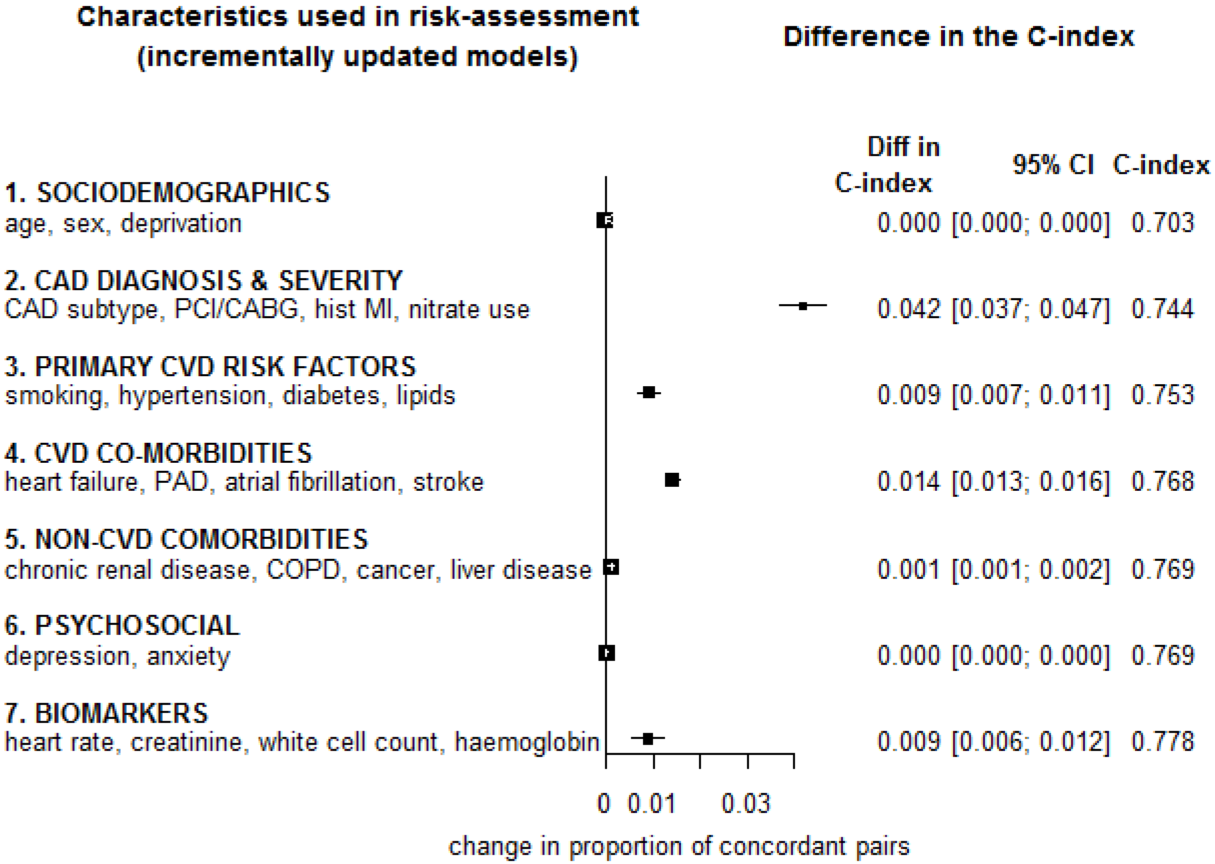


A ‘gold standard’ for estimating relevant risks?

Rapsomaniki et al, CALIBER, EHJ 2014;35(13):844-52

Prognostic models using linked EHR:

Which *clinically recorded* factors add to discrimination?



Characteristics used in risk-assessment (incrementally updated models)	Diff in C-index	95% CI	C-index
1. SOCIODEMOGRAPHICS age, sex, deprivation	0.000	[0.000; 0.000]	0.703
2. CAD DIAGNOSIS & SEVERITY CAD subtype, PCI/CABG, hist MI, nitrate use	0.042	[0.037; 0.047]	0.744
3. PRIMARY CVD RISK FACTORS smoking, hypertension, diabetes, lipids	0.009	[0.007; 0.011]	0.753
4. CVD CO-MORBIDITIES heart failure, PAD, atrial fibrillation, stroke	0.014	[0.013; 0.016]	0.768
5. NON-CVD COMORBIDITIES chronic renal disease, COPD, cancer, liver disease	0.001	[0.001; 0.002]	0.769
6. PSYCHOSOCIAL depression, anxiety	0.000	[0.000; 0.000]	0.769
7. BIOMARKERS heart rate, creatinine, white cell count, haemoglobin	0.009	[0.006; 0.012]	0.778

Origin of data is EHR therefore implementation of risk prediction models in decision support tools (with evaluation) is feasible

Rapsomaniki et al, CALIBER, EHJ 2014;35(13):844-52

Farr
The Farr Institute of
Health Informatics Research

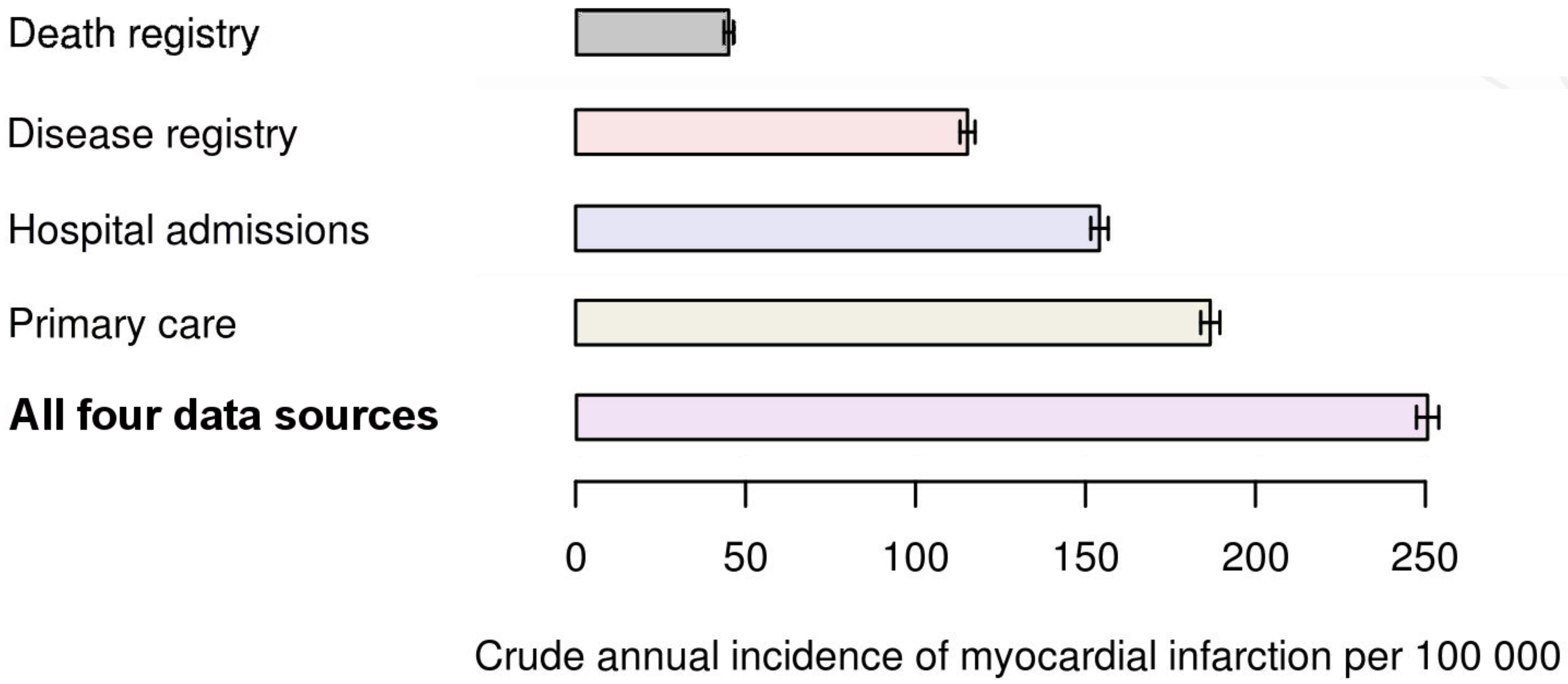
Discovery

Trials

Outcomes research/real world evidence

Public Health

Outcomes assessment: importance of linking multiple record sources

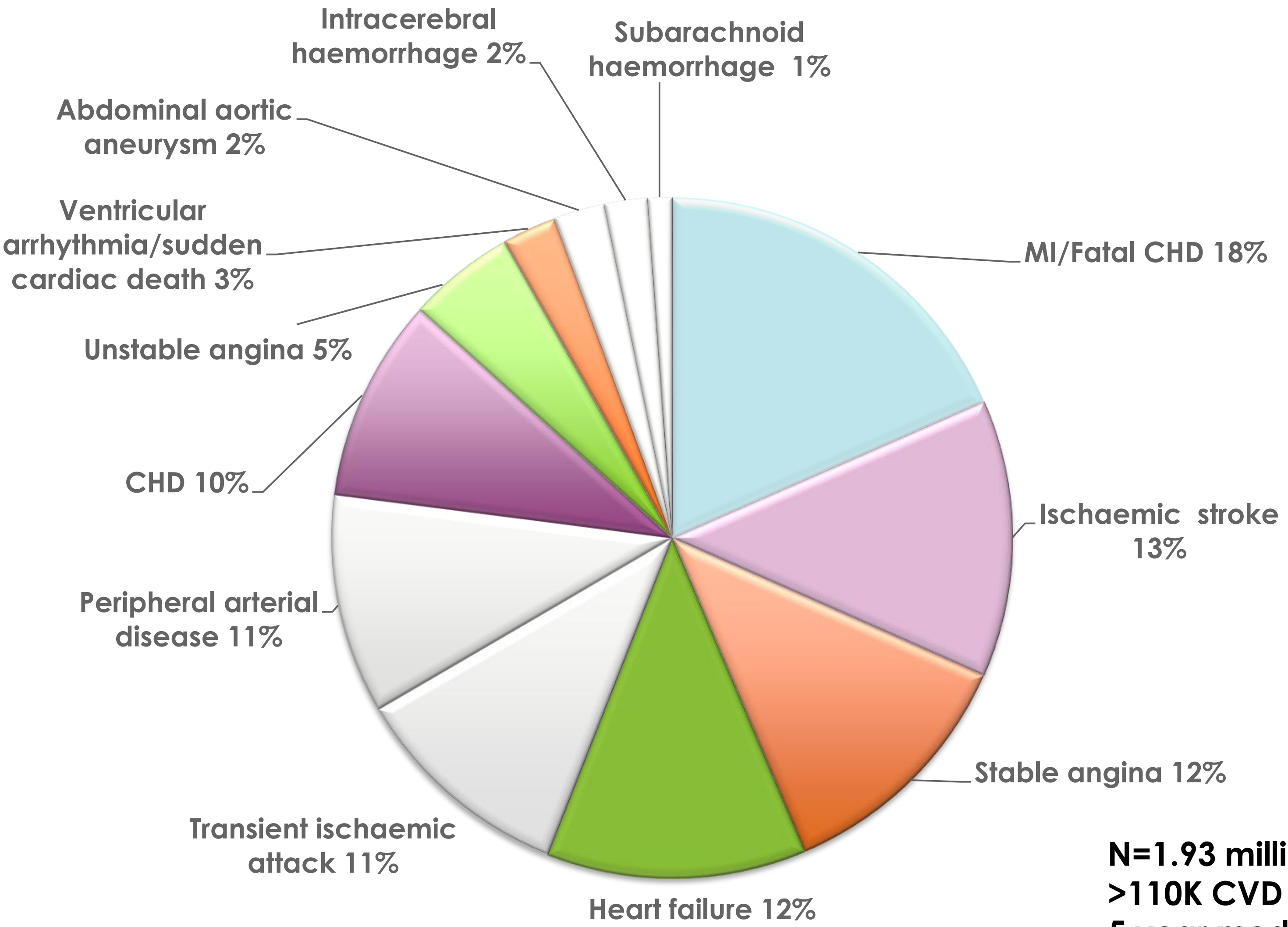


Herrett et al, CALIBER, BMJ 2013;346:f2350

The Farr Institute of Health Informatics Research

How does CVD first present?

In the real world, today



N=1.93 million patients
>110K CVD events
5 year median follow-up

CALIBER 2014, under review

Discovery

Trials

Outcomes research/real world evidence

Public Health

What is the role of the Farr Institute?

The Farr Institute of
Health Informatics Research

Drought

- **Data**
  - Need much wider national record linkages – CPRD-NICOR-HES
  - Need to liberate 'submerged' deeper hospital phenotypes
  - Need to converge EHR, omics and imaging

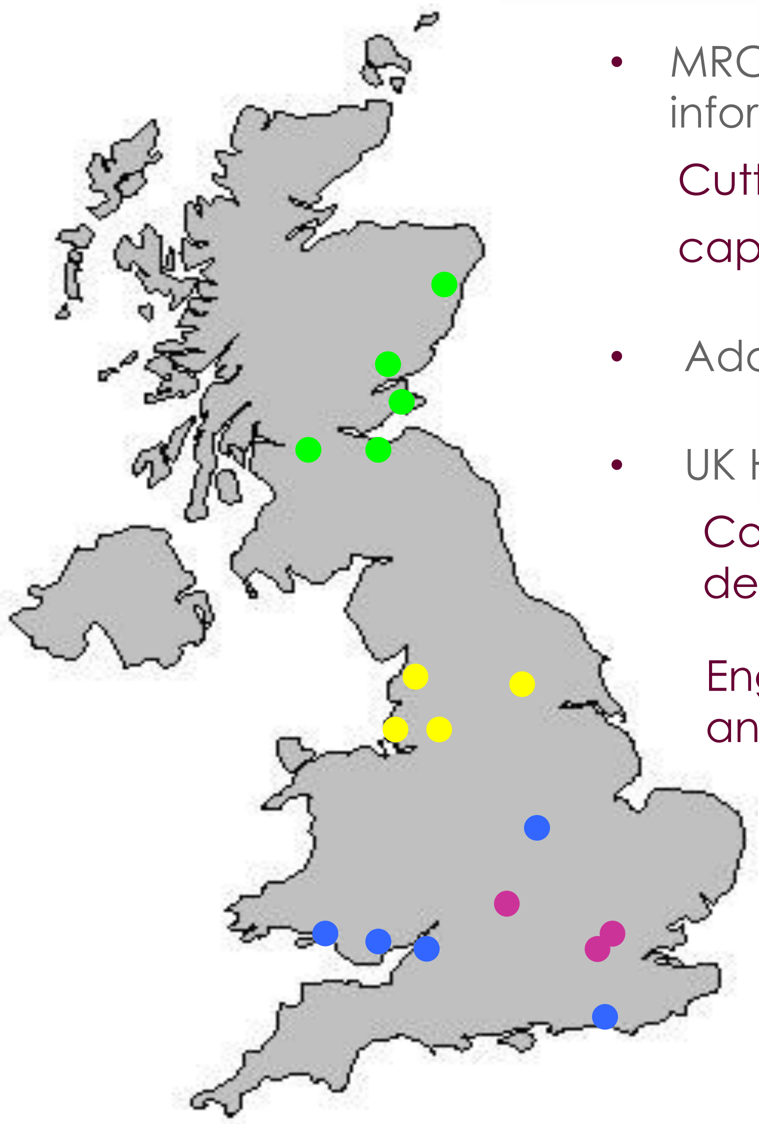
- **Tools**
  - Health informatics '20 years behind bioinformatics'
  - And UK 20 yrs behind US?

- **People**
  - (re) building **public trust** (care.data)
  - Not nearly enough **clinicians** with the training and opportunity to drive improvements in care (and research) through data (cf new US sub-specialty)
  - Careers for technical staff
  - **Interdisciplinarity**

The Farr Institute of Health Informatics Research

Strengthening health informatics research

wellcome trust
DH Department of Health
British Heart Foundation
MRC Medical Research Council
Llywodraeth Cynulliad Cymru Welsh Assembly Government
ESRC Economic & Social Research Council
Cancer Research UK
Chief Scientist Office
Arthritis Research UK Providing answers today and tomorrow
EPSRC Engineering and Physical Sciences Research Council



- MRC coordinated 10-partner **£19m** call for e-health informatics research centres across the UK

  Cutting edge research using data linkage

  capacity building

- Additional **£20m** capital to create Farr Institute

- UK Health Informatics Research Network

  Coordinate training, share good practice and develop methodologies

  Engage with the public, collaborate with industry and the NHS

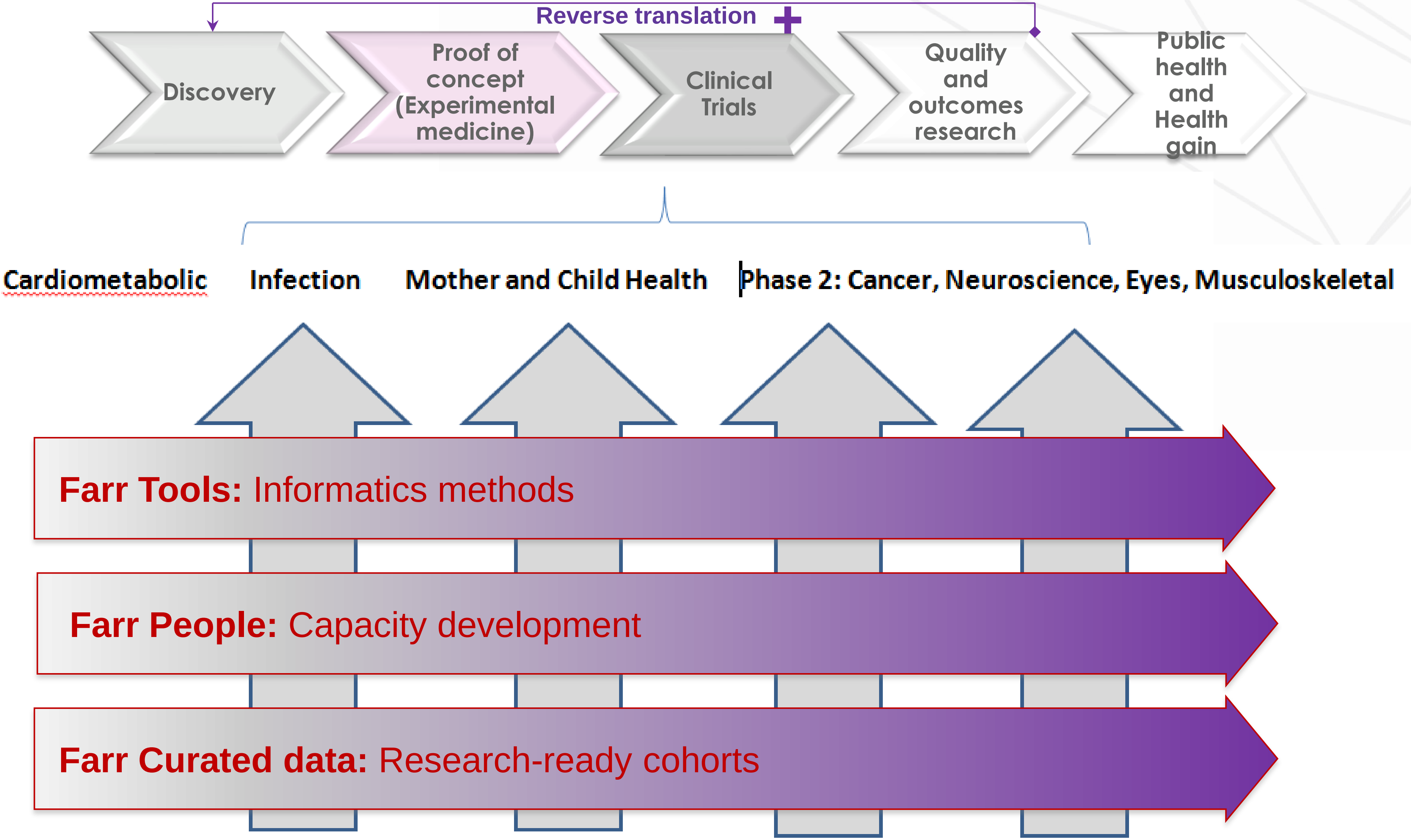
● **Farr London**

● **Farr Scotland**

● **Farr at Swansea, Wales**

● **Farr N8, Manchester**

What are the aims?
= research along the translational pathway



Reverse translation

Discovery

Proof of concept (Experimental medicine)

Clinical Trials

Quality and outcomes research

Public health and Health gain

Cardiometabolic

Infection

Mother and Child Health

Phase 2: Cancer, Neuroscience, Eyes, Musculoskeletal

Farr Tools: Informatics methods

Farr People: Capacity development

Farr Curated data: Research-ready cohorts

Rapid evolution of initiatives: emphasis on infrastructure

Farr
The Farr Institute of Health Informatics Research

JULY 2013

Genomics england

JULY 2013

Administrative Data Research Centre England

OCTOBER 2013

NHS
National Institute for Health Research

November 2013
Health Informatics Collaboration
(sharing hospital data across 5 Biomedical Research Centres)

MRC Medical Research Council

February 2014
Medical Bioinformatics Awards

UCLPartners
Academic Health Science Partnership

Who was William Farr?

"Diseases are more easily prevented than cured and the first step to their prevention is the discovery of their exciting causes."

1807-1883

Compiler of Scientific Abstracts at General Register Office
…aka 'Big data health'

Gave us cause of death and International Classification of Disease

Local actions e.g. Victoria Park

Conclusion

- Most of what we know about mortality and morbidity has come from much ‘smaller’ data than is currently available to researchers

- Personalisation is a secular phenomenon across multiple sectors in society: Medicine offers vanguard and laggard examples!

- If informatics is about data, tools and people – then it is the people which need most urgent development.

Farr London (original) Investigators

CARDIOVASCULAR

- **Mike Barnes**, Director of Bioinformatics
- **James Carpenter**, Professor of Medical Statistics
- **John Deanfield**, Professor of Paediatric Cardiology
- **Mark Caulfield**, Professor Clinical Pharmacology
- **Spiros Denaxas**, Health Informatics Senior Research Associate
- **Nicholas Freemantle**, Professor of Clinical Epidemiol and Biostatistics
- **Harry Hemingway**, Professor of Clinical Epidemiology
- **Aroon Hingorani**, Professor of Genetic Epidemiology
- **Steffen Petersen**, Reader in Advanced Cardiovascular Imaging
- **John Robson**, GP, Clinical lead for the Clinical Effectiveness Group
- **Liam Smeeth**, Professor of Epidemiology
- **Adam Timmis**, Professor of Clinical Cardiology

INFORMATICS

- **Anne Blandford**, Professor of Human–Computer Interaction
- **Peter Coveney**, Professor of Physical Chemistry
- **James Freed,** Head of Health Intelligence and Standards
- **Dipak Kalra**, Professor of Health Informatics
- **John Shawe-Taylor**, Professor of Computing
- **Paul Taylor,** Reader in Health Informatics
- **Alan Wilson**, Professor of Urban Regional Systems

MOTHER & CHILD

- **Peter Brocklehurst**, Professor of Women's Health
- **Tito Castillo,** Chief Operating Officer, LIFE Study
- **Carol Dezateux**, Professor of Paediatric Epidemiology
- **Ruth Gilbert**, Professor of Clinical Epidemiology
- **Irene Petersen**, Senior Lecturer Epidemiology and Medical Statistics
- **Judith Stephenson**, Professor of Reproductive and Sexual Health
- **Phil Koczan,** Chief Clinical Information Officer
- **Irwin Nazareth**, Professor of Primary Care and Population Science
- **Max Parmar**, Director of MRC Clinical Trials Unit

INFECTION

- **Mike Catchpole**, Head of Epidemiology and Surveillance
- **Andrew Hayward**, Senior Clinical Lecturer in Infection
- **Richard Pebody**, Head of the Seroepidemiology Programme
- **Deenan Pillay**, Professor of Virology

PHASE 2 CLINICAL WORKSTREAMS

- **Andy Goldberg**, Senior Lecturer in Trauma and Orthopaedics
- **Anthony Moore**, Professor of Ophthalmology
- **Kathy Pritchard-Jones,** Professor of Paediatric Oncology
- **Martin Rossor**, Professor of Neurology & Director of DeNDRON