How do we do health data science that delivers impact?

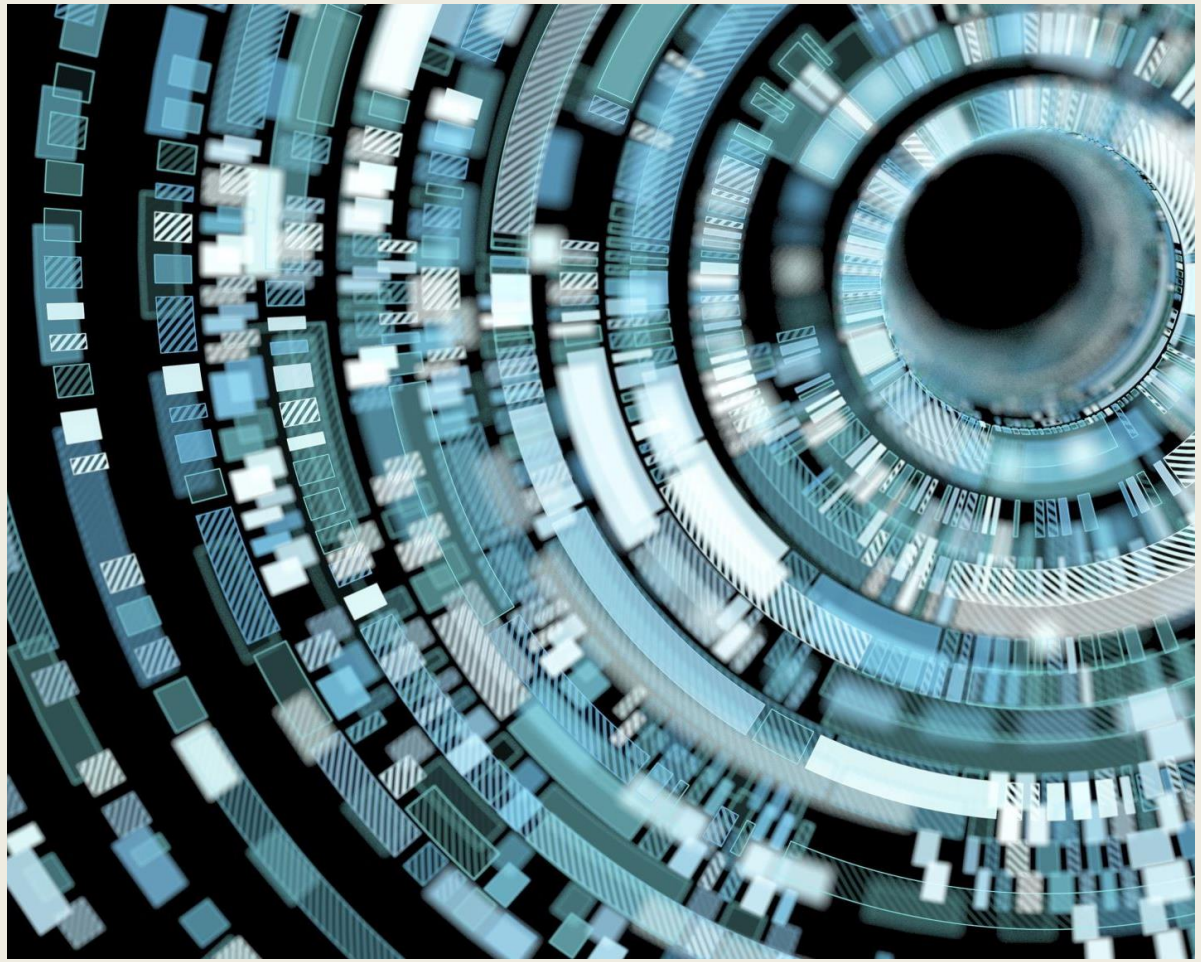
ARC KSS Research Week 2024

Dr Elizabeth Ford, Reader in Health Data Science
Brighton and Sussex Medical School

brighton and sussex
medical school

Content Overview and Learning Aims

1. Understanding what impact means
2. What is health data science?
3. What is routinely collected health and care data?
4. Other sources of health “big data”
5. Types of health research questions we can ask of “big data”
6. Building projects and teams that deliver impact – case study



What is research impact ?

NIHR

- Research having real world effects
- Making a meaningful difference in people's lives
- Contribute to improving health and care for the population

Health Data Research-UK

Research produces demonstrable positive effects on:

- Data science methods and technology
- Public trust
- Health and well-being
- Decision-making
- Research culture and capacity

How does impact from research happen?

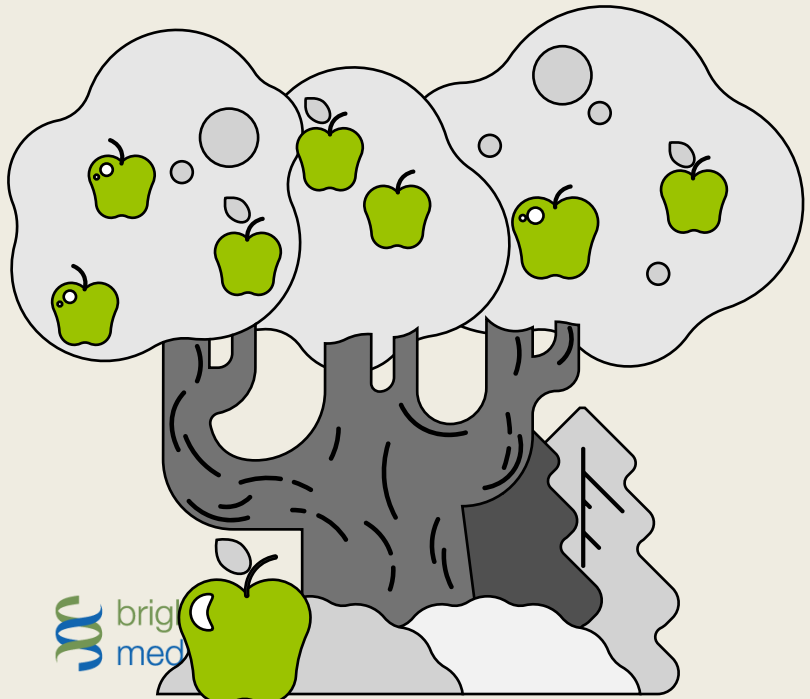
New knowledge needs to be

Generated

Contextualised

Put into Practice

Who needs to be involved to make this happen?



Stakeholders…

Having a “stake” in something…

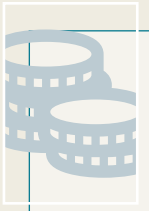
…means having an **interest, investment,** or **involvement** in something, or that something is **important** to you

Major Healthcare Stakeholders – 4Ps

Patients

Providers

Policymakers



Payors

Who will be your end-users or knowledge users?

It is never too early to involve them!

What do we mean by applied health research?

Applied health research aims to address the immediate issues facing the health and social care system.

It involves research on human, individual, and population health issues that is intended to have an impact on:

- The health of individuals
- The health of populations and communities
- Decisions about government health policy
- Health system organization
- Healthcare delivery

It looks to find practical solutions for existing problems.

It involves a wide range of disciplines and can be multi-disciplinary and collaborative.

What is Health Data Science?

Health data science

- **The application of data science methods to healthcare data**
- This involves collecting, cleaning, analysing, and interpreting large datasets

What are the most common aims?

- Disease diagnosis and prognosis
- Drug discovery and development
- Personalised Medicine
- Improving efficiency of healthcare systems
- Addressing public and population health challenges such as epidemics, health disparities, and chronic diseases

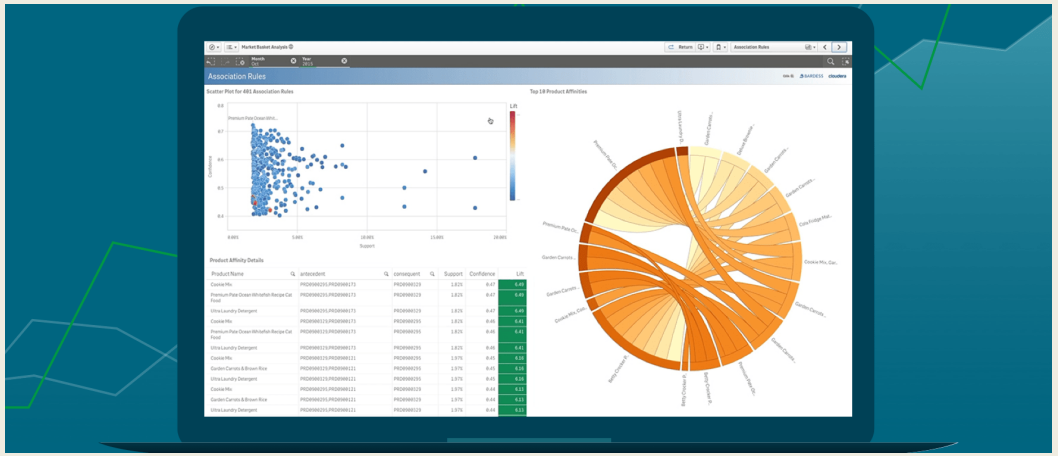
Tools and techniques commonly used in health data science

Statistical analysis: Descriptive statistics, hypothesis testing, regression analysis, and survival analysis.

Machine learning: Classification, regression, clustering, and deep learning.

Natural language processing: Extracting information from unstructured text data, such as medical notes, reports and letters

Data visualization: Creating visual representations of data to communicate findings effectively reports and clinical notes

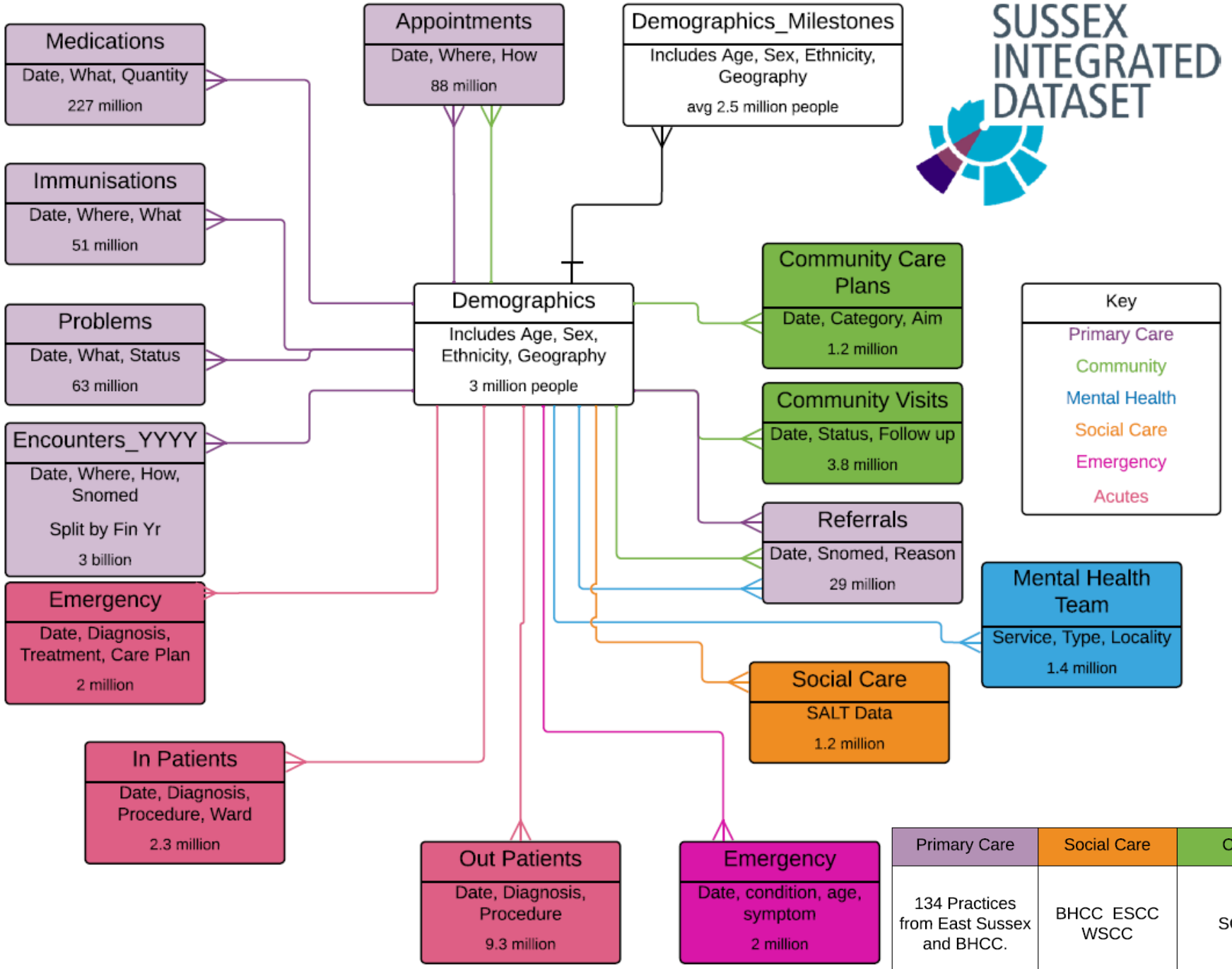


What do we mean by routinely collected health and care data?

Data collected in the course of interactions with the health service

This could include:

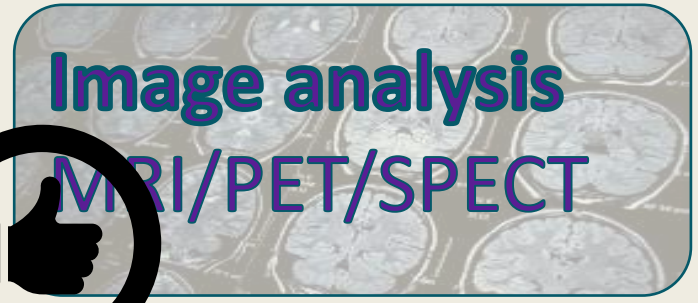
- GP clinic notes (unstructured), prescriptions, and coded data (Read or Snomed Codes)
- Hospital coded data – ICD codes – recorded for reimbursement for episodes of care:
  - Emergency Department
  - Outpatient Care
  - Inpatient Care
  - Discharge summaries (unstructured)
  - Scan results, lab tests and pathology reports
- Mental health patient notes (unstructured) and some coded data
- Nursing records
- Disease registries e.g. cancer, MI, infectious diseases.
- Social care data (on provision of SC by local authorities)

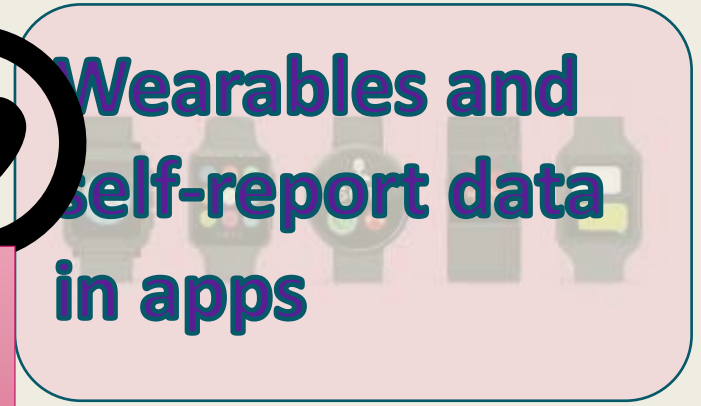


Primary Care	Social Care	Community	Mental Health	Acutes	Emergency
134 Practices from East Sussex and BHCC.	BHCC ESCC WSCC	SCFT ESHT	SPFT	UHSx, ESHT, QVH	SECAMB

Other sources of health big data

Other sources of health data

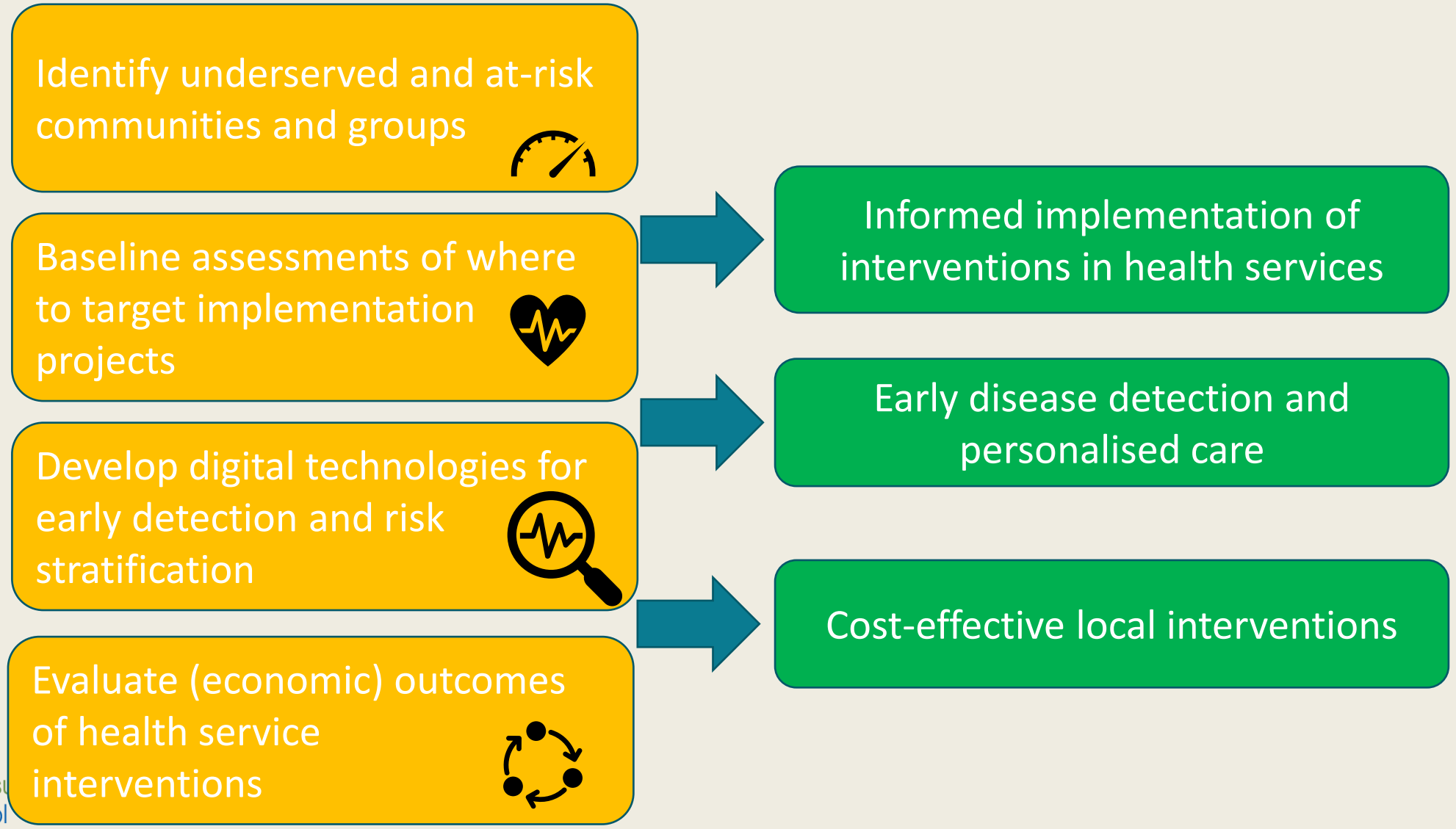




Combination of multi-modal data

brighton and sussex medical school

What do we aim to do locally with routinely collected, linked datasets?



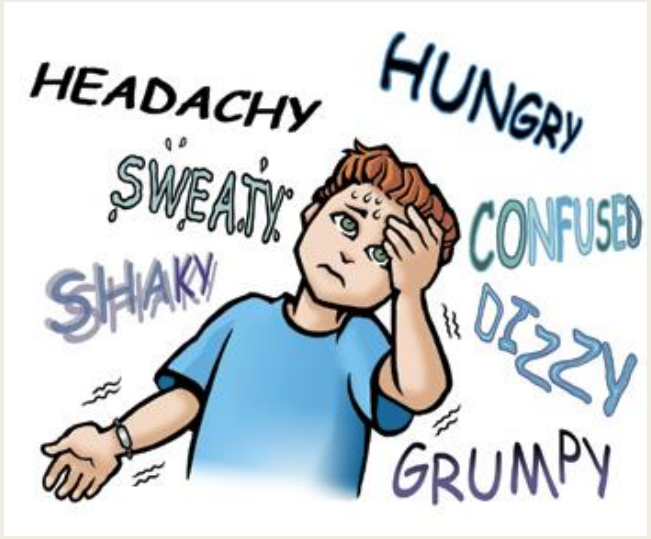
What health questions can we ask and answer with health data?

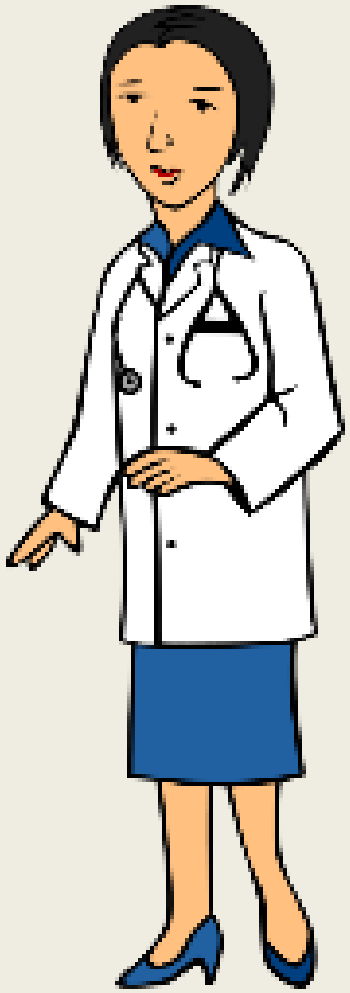
Need to form research question AROUND the data

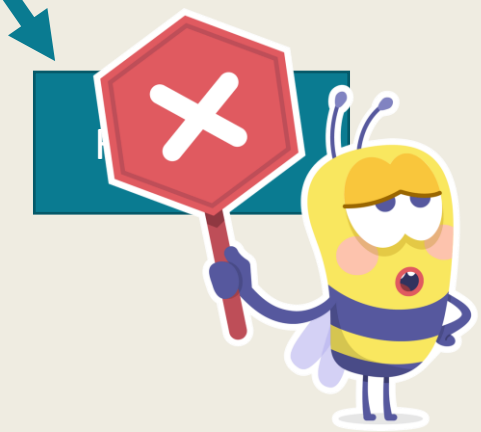
We are stuck with what is there, so we can only ask questions if the data has been collected, and we know it is reasonable quality.

Electronic Health Record data is recorded by the clinician.

So having information about someone's condition depends on the patient telling the clinician and the clinician recording it in a code in the record.


HEADACHY
HUNGRY
SWEATY
CONFUSED
SHAKY
DIZZY
GRUMPY





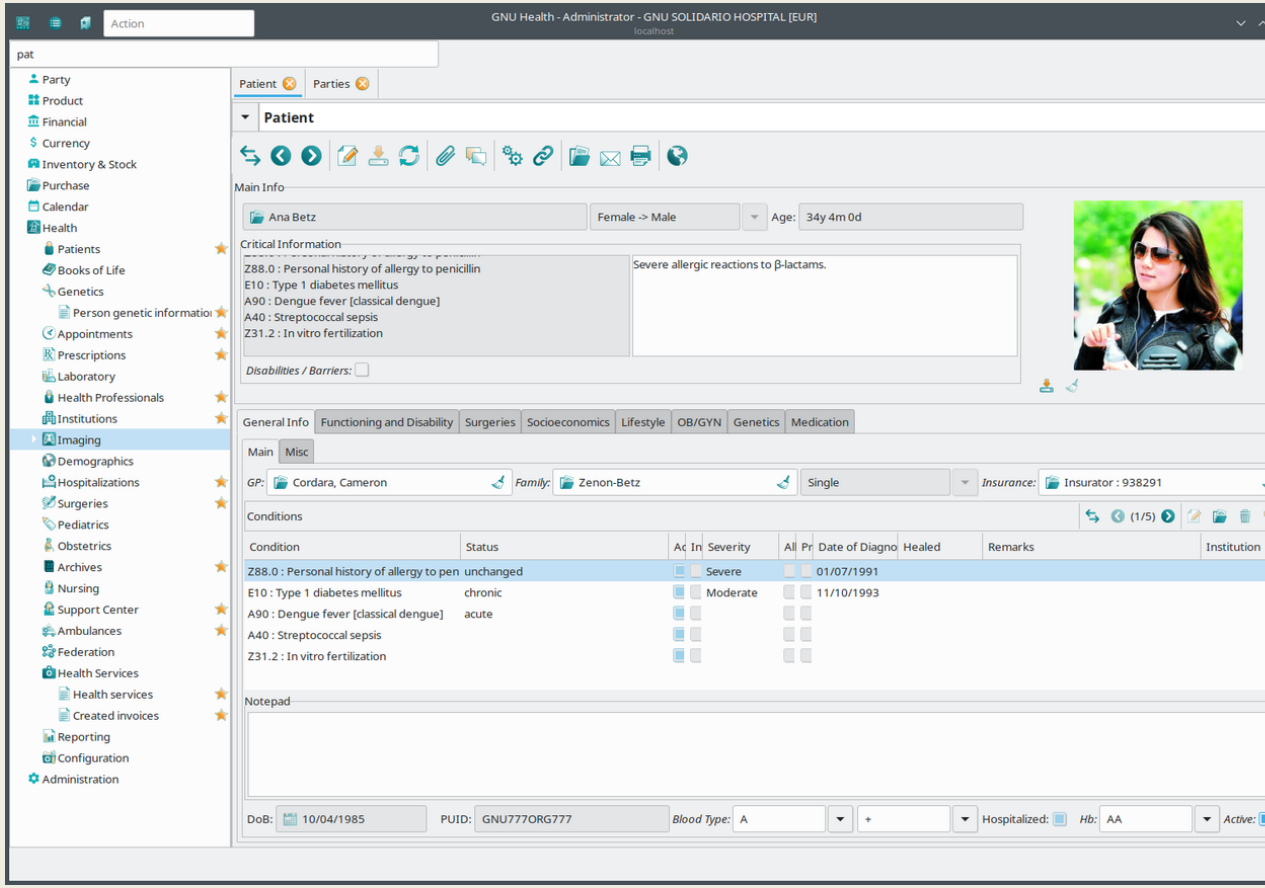
Example: Men's experience of urinary problems and erectile dysfunction after prostate cancer treatment

What might affect the numbers and characteristics of men we find with these symptoms if we use anonymised GP data?

Type in the chat…

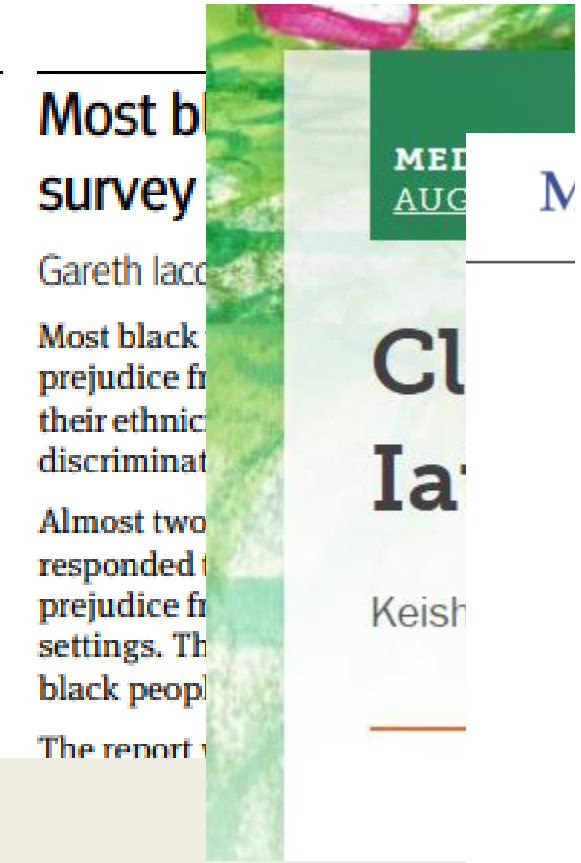
Constraints on what data gets entered

- **Coding structure** – flexibility and limits
- **Patient record software interface** – ease of use
- **Time available** for documentation
- **Motivation for documentation** – billing / record for own memory / inform next clinician / medico-legal safety netting?
- **Clinical reasoning and filtering of information** – what is most important?
- **Consideration (or lack thereof) of secondary purposes**



brighton and sussex medical school

Structural biases in healthcare will be *baked into* data



ARTICLE JAN 18, 2018

Discrimination Prevents LGBTQ

Disability Bias in Health Care

Review > J Community Health Nurs. 2018 Jan-Mar;35(1):28-37.

doi: 10.1080/07370016.2018.1404832.

Barriers to Health Care Access for Low Income Families: A Review of Literature

Malerie Lazar [1], Lisa Davenport [1]

How to form a good research question

A concise question is very important.

Why?

If you receive 3 billion rows of data, what are you going to look at?

You will need to define all your variables using lists of clinical codes

Carefully plan in advance your exposures, confounders and outcomes.

E.g. consider the following:

Which mental health symptoms are most associated with referrals to secondary care mental health services?

- Which symptoms will you choose?
- What about physical symptoms? Tummy aches, fast heart rate, neck/jaw stiffness
- How would you know if they are mild or severe?

I want to make a predictive model!

Questions to ask before you start:

What do you want to predict?

Are there good quality data representing that outcome?

Does any end-user want the prediction?

Is there any point in making a prediction?

Are the useful or important predictors well captured?

Will it help workflow or add extra pressure to the system?

If someone is flag-positive in the model, do you have the resources to care for them or change their outcome?

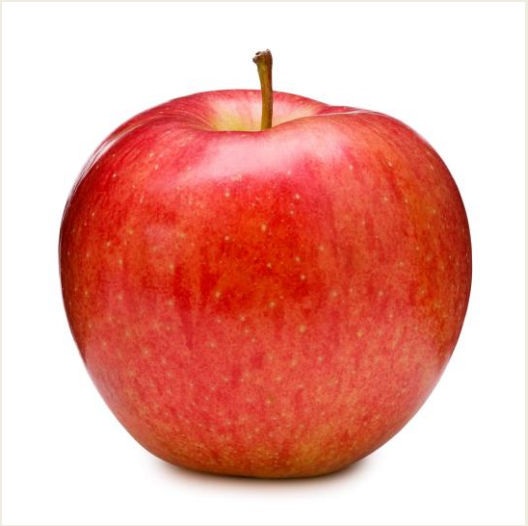
What is the likelihood of an accurate prediction? (beware of false positive outcomes…)

Create your project plan in a team

- Clinicians must help the analyst understand the clinical workflow
- What are the drivers of different decisions, referrals, treatments or recording?
- Ask front-line clinicians to help with code lists and variable definitions
- What is clinical information likely to be recorded vs social information unlikely to be captured?
- How “granular” is the data recording in the service you are looking at?
    - E.g. reasons for A&E attendance.

More importantly:

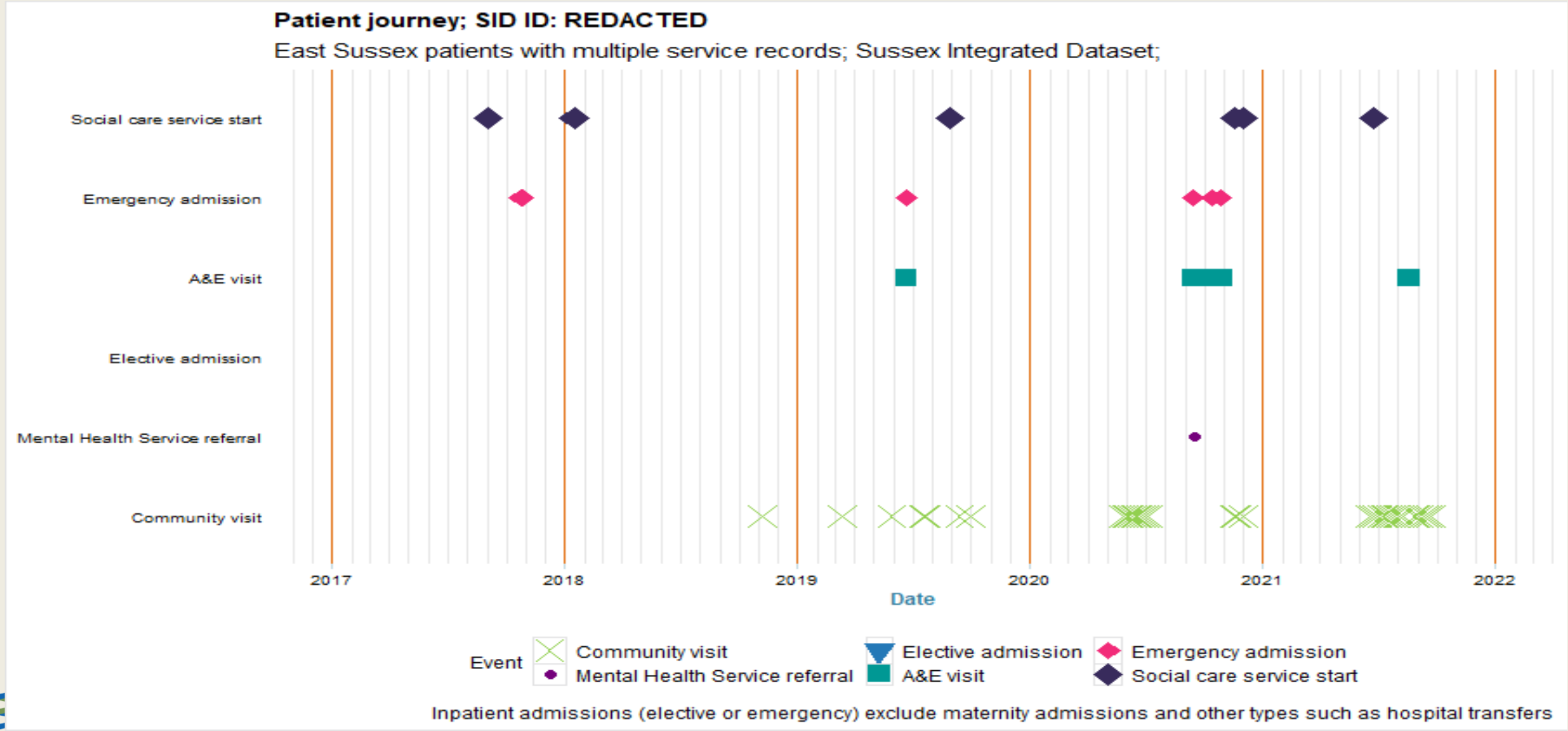
- Have frontline clinicians and commissioners **determine the questions you ask**
- Then you will generate knowledge most likely to be actionable.



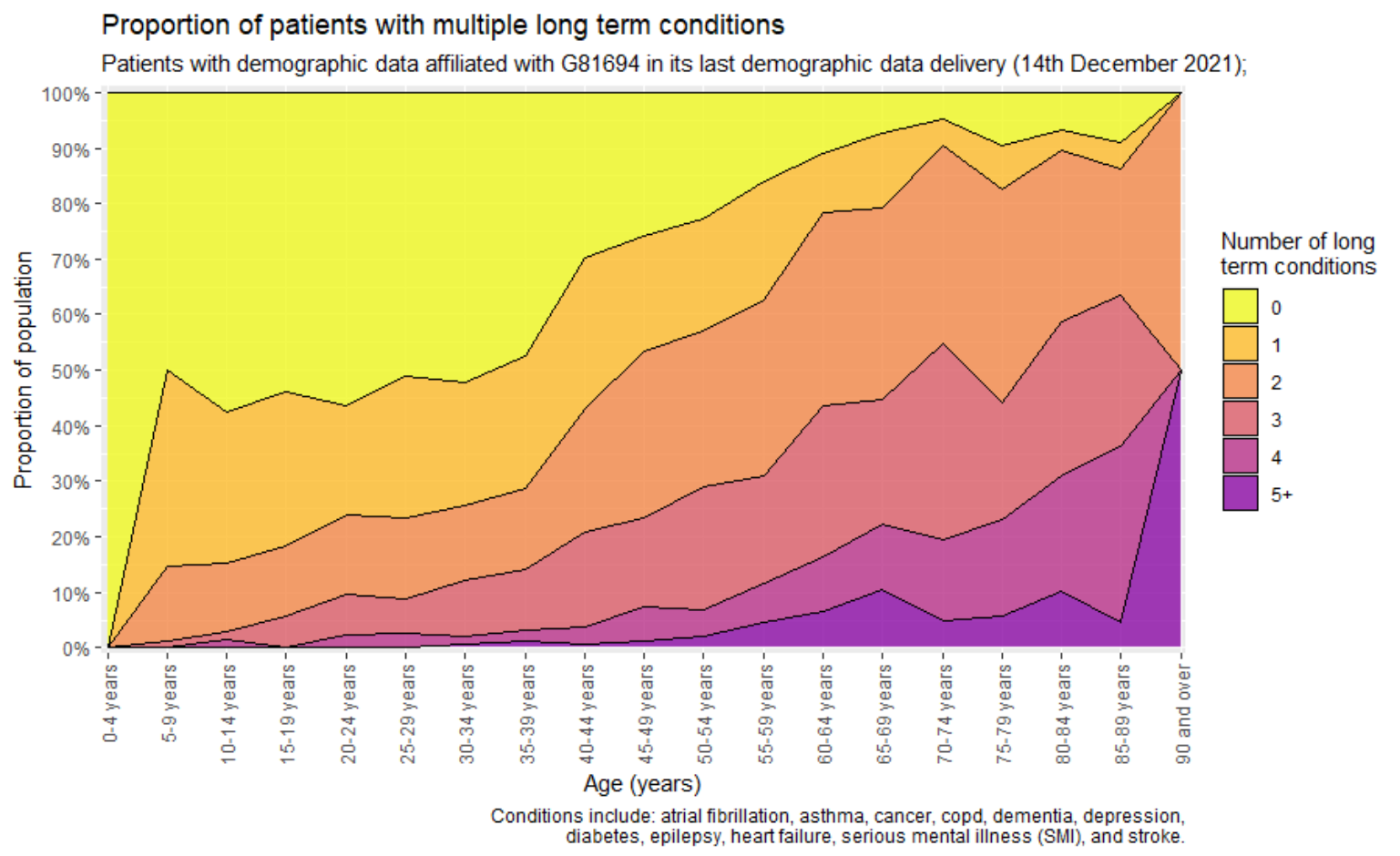
Using data to address health inequalities

brighton and sussex medical school

Patient theographs can track a patient's care through multiple providers… and spot where they fall through gaps



We can examine the rates of multiple long term conditions by GP practice to examine excess comorbidity by location



Identifying number of comorbidities by age, stratified by GP practice

- Can locate area with higher burden of multi-morbidity or earlier onset

brighton and sussex medical school

Case study: Predicting dementia

ASTRODEM

Using astrophysics to close the diagnosis gap for dementia

Elizabeth Ford

ASTRODEM

brighton and sussex medical school

wellcome

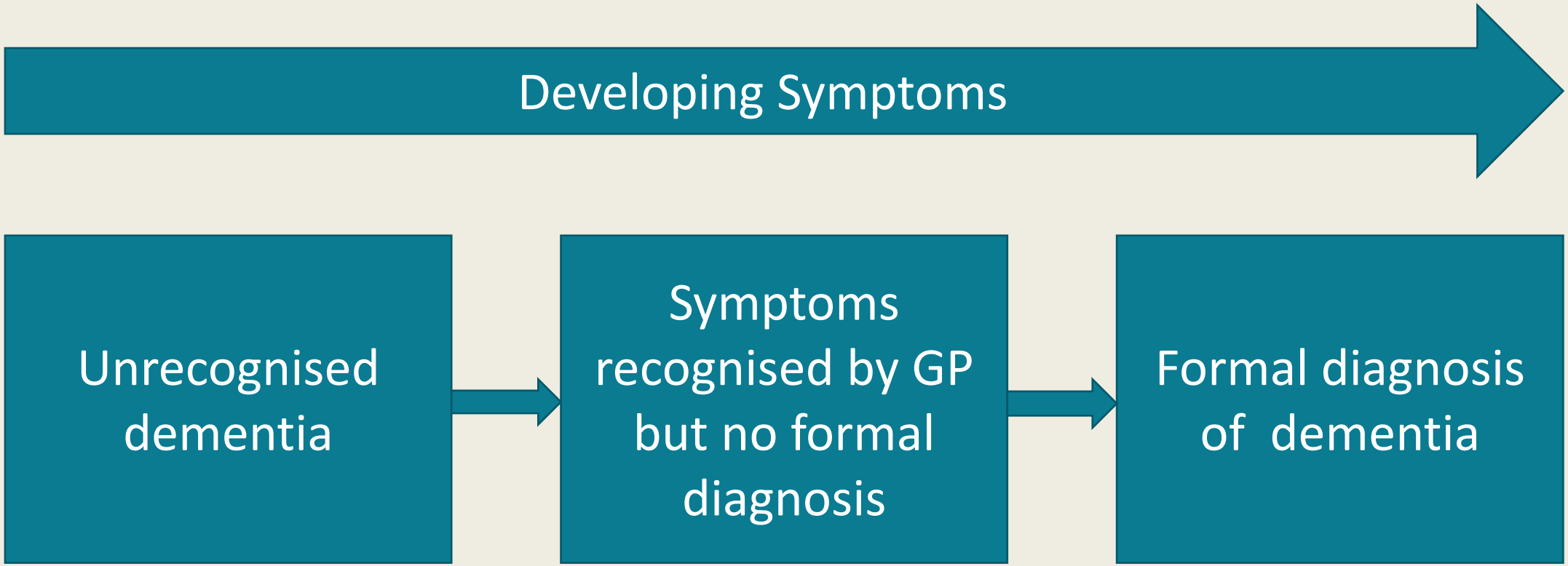
Early Detection of Dementia – Why?

Currently, diagnosis is often made:

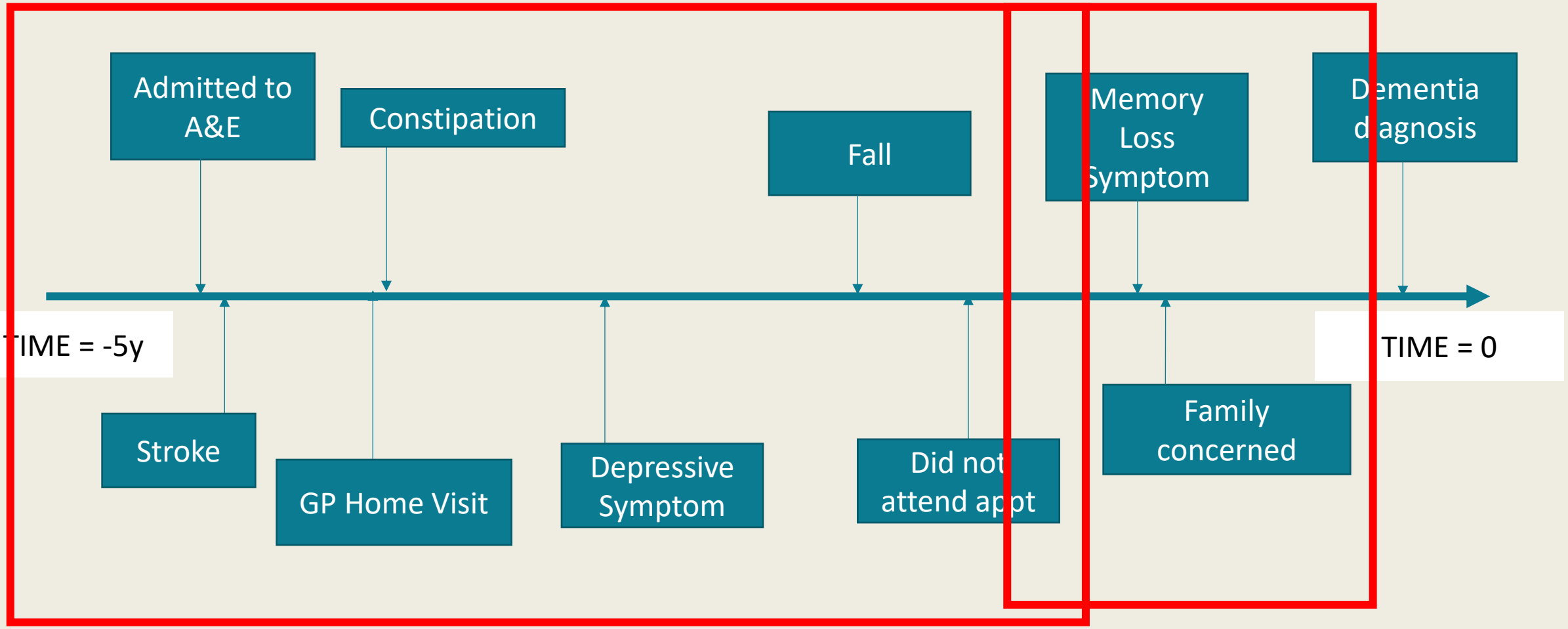
- At a time of crisis (e.g. following a burn or a fall)
- When the family are struggling to cope with care needs
- Too late for the person with dementia to express their financial and care preferences and make plans.
- An average of 3.5 years after patient meets diagnostic criteria.

Three reasons to detect dementia earlier

- To allow patients time to plan, express their wishes, and organise care
- To enable access to disease modifying treatments (when available)
- For recruitment to clinical trials


Developing Symptoms
Unrecognised dementia
Symptoms recognised by GP but no formal diagnosis
Formal diagnosis of dementia

A primary care clinical record in the 5 years before dementia diagnosis



Machine learning models

- Case-control study – 50% with dementia, 50% controls,
- GP records for 5 years prior to diagnosis in 95,000 people
- Baseline method: Logistic Regression with LASSO penalisation

Can we improve on this using:

Neural Networks

Support Vector Machine

Random Forest

Naïve Bayes

Methods for assessing models

Area under ROC

Regression Coefficients

brighton and sussex medical school

Typical predictors in the models

Prodromal symptoms:
- Behaviour change
- Disorientation or wandering
- Personality change

Prescriptions:
- Antidepressants
- Antipsychotics

Indicators of reduced coping:
- Self-neglect
- Difficulty managing
- Missed appointments
- Visits to A&E
- Falls

Historical factors:
- Previous diagnosis of schizophrenia or bipolar disorder
- Family history of dementia

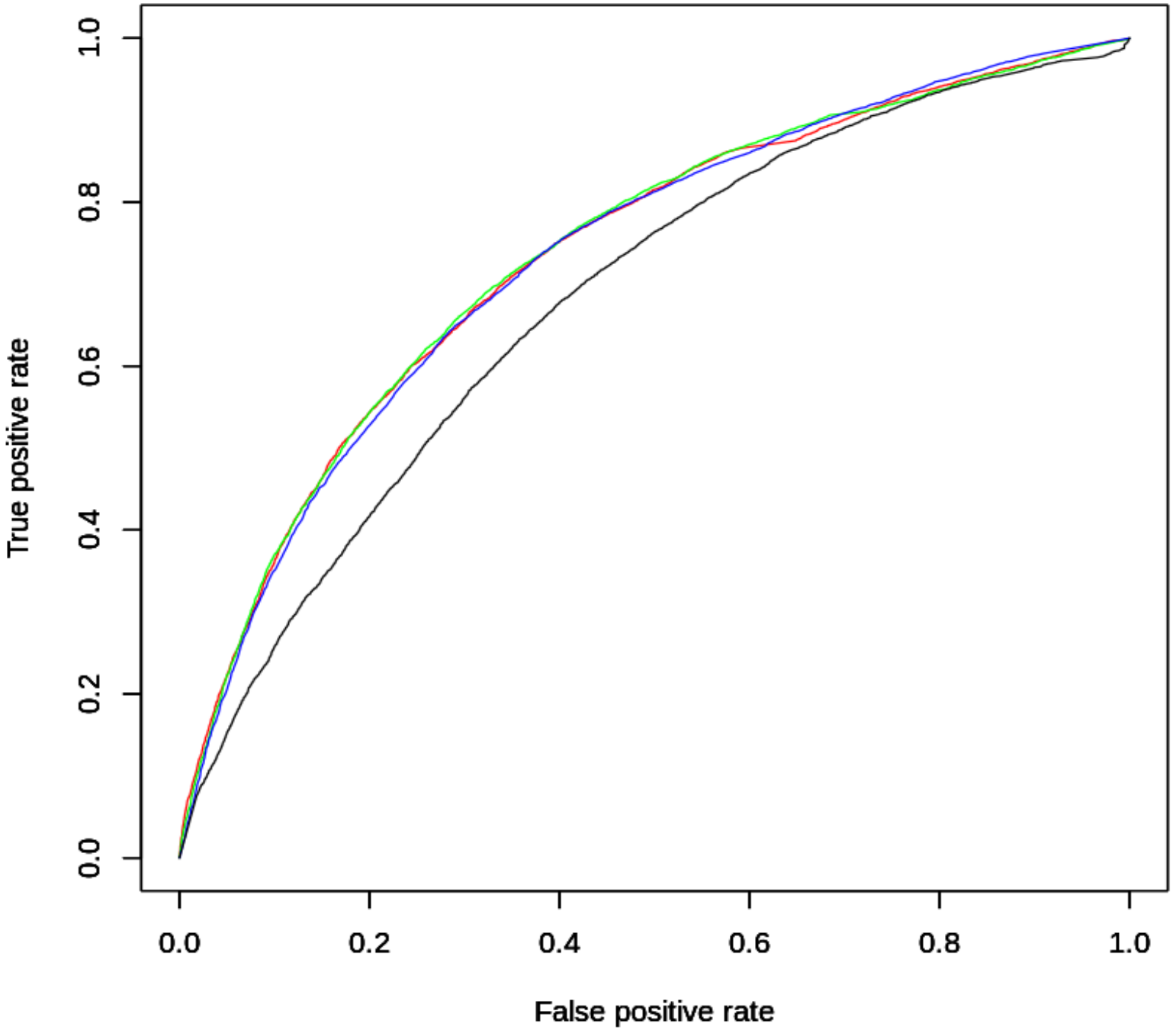
Social factors:
- Family Concerned
- GP home visit
- Social services
- Receiving care

Other conditions:
- Cerebrovascular Disease/Stroke
- Depression
- Diabetes
- Epilepsy

Receiver Operating Characteristic Curve

Model	AUC
Log reg	0.74
Neural Network	0.74
SVM	0.74
Random Forest	0.73
Naïve Bayes	0.68

Red = logistic regression; Green = SVM;
Blue = Random Forest; Black = Naïve Bayes



brighton and sussex medical school

Project 2: DemRisk

Funded by:

brighton and sussex
medical school

Identifying patients at high risk for dementia in next 5 years

Cohort study – unselected older people followed for 5 years to see if they got a new/first dementia diagnosis.

Using GP patient data on > 1 million people.

60 risk factors evaluated

Age 60-79

Discrimination and Calibration were good (**Harrell's C 0.78** (95% CI: 0.78 to 0.79)).

37% of patients in the top 1% of risk scores received a dementia diagnosis within 5 years.

Age 80-89 Model discrimination was lower, but prevalence was higher.

79% of those in the top 1% of risk scores subsequently developed dementia.

But is prediction/detection of dementia a good idea?

Received: 16 September 2022 | Revised: 12 January 2023 | Accepted: 13 January 2023

DOI: 10.1002/widm.1492

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

ADVANCED REVIEW

Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia

Elizabeth Ford[1] | **Richard Milne**[2,3] | **Keegan Curlewis**[4]

[1]Department of Primary Care and Public Health, Brighton and Sussex Medical School, Brighton, UK

[2]Kavli Centre for Ethics, Science and the Public, University of Cambridge, Cambridge, UK

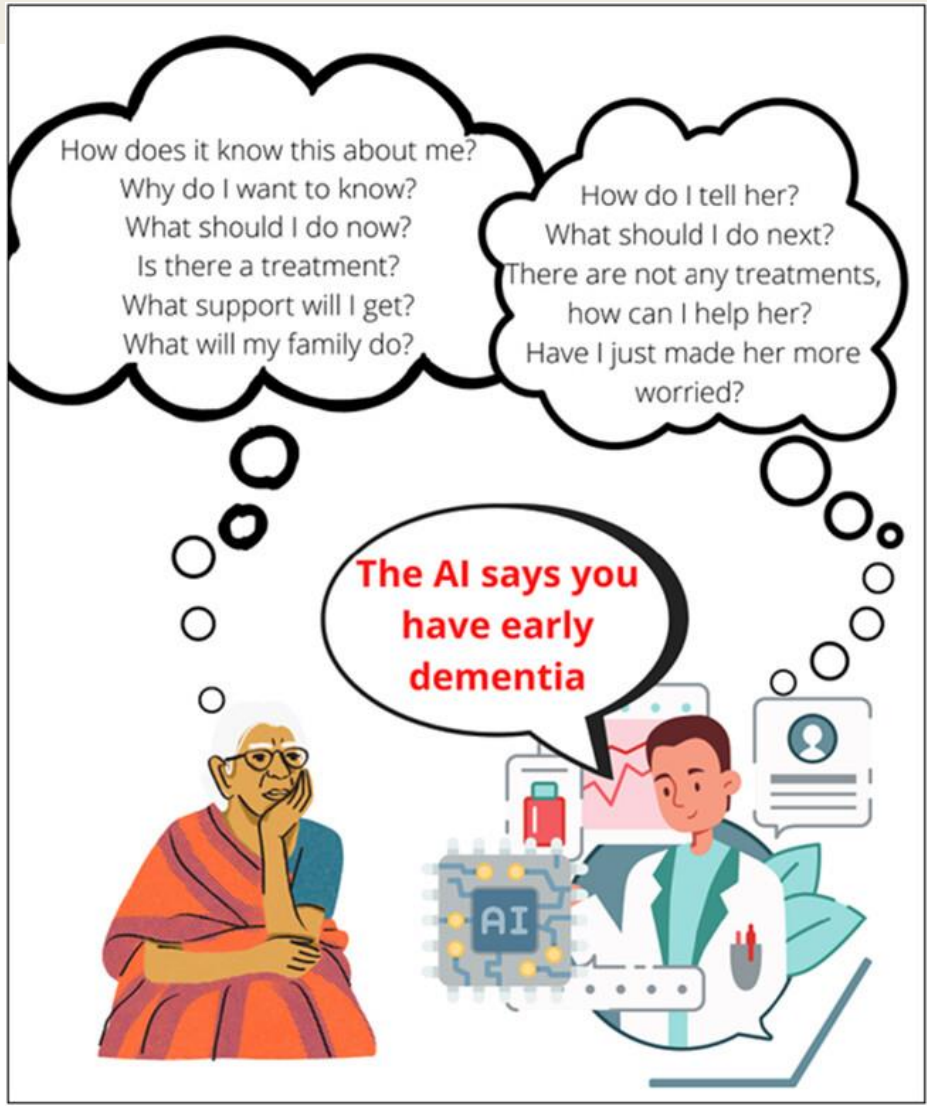
[3]Engagement and Society, Wellcome Connecting Science, Cambridge, UK

[4]Royal Free London NHS Foundation Trust, London, UK

Correspondence
Elizabeth Ford, Department of Primary Care and Public Health, Brighton and

Abstract

Dementia poses a growing challenge for health services but remains stigmatized and under-recognized. Digital technologies to aid the earlier detection of dementia are approaching market. These include traditional cognitive screening tools presented on mobile devices, smartphone native applications, passive data collection from wearable, in-home and in-car sensors, as well as machine learning techniques applied to clinic and imaging data. It has been suggested that earlier detection and diagnosis may help patients plan for their future, achieve a better quality of life, and access clinical trials and possible future disease modifying treatments. In this review, we explore whether digital tools for the early detection



What sort of clinical decision support do GPs want?

Ford *et al. BMC Med Inform Decis Mak* (2021) 21:193
https://doi.org/10.1186/s12911-021-01557-z

BMC Medical Informatics and Decision Making

RESEARCH Open Access

Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners

Elizabeth Ford[1*], Natalie Edelman[1,2], Laura Somers[1], Duncan Shrewsbury[1], Marcela Lopez Levy[3], Harm van Marwijk[1], Vasa Curcin[4] and Talya Porat[5]

Dementia Risk Prediction - Views of Key Stakeholders

Semi-structured Interviews with:

36 Primary Care Practice Staff (Sussex, Greater Manchester, Newcastle)

48 Primary Care Patients (age 40-79, without dementia).

-How do elderly patients and carers feel about being offered a routine risk assessment by their GP to assess their chance of developing dementia

-What do they consider to be the main benefits and drawbacks of such a tool?

-How should the results of the tool best be communicated? What should happen next?

NIHR | National Institute for Health and Care Research

brighton and sussex medical school

Results from healthcare professionals

HCP results:

1. **Yes, but only if there is an intervention**
2. **Yes, it could help people to find out about modifiable risk factors and change their lifestyles**
3. **You can't do anything about genetic risk factors, so it's much better to look at modifiable risk factors than anything else**
4. **Some people might be able to turn their risk around**
5. **It would be helpful to have a risk assessment tool for routine care. It could lead to earlier diagnosis and possibly some good treatment that could delay the onset of dementia.**
6. **It could help families because then they would be prepared.**

Results from Patients

“Yes, I would want to know …… but some people wouldn’t.”

“I’m sure that some people would be very anxious about it and may be a bit reticent because they might not want to find out. I think it should be a personal choice whether people want to have it done or not”

“**It’s helpful to know your risk so you can take more control over your diet and lifestyle**.”

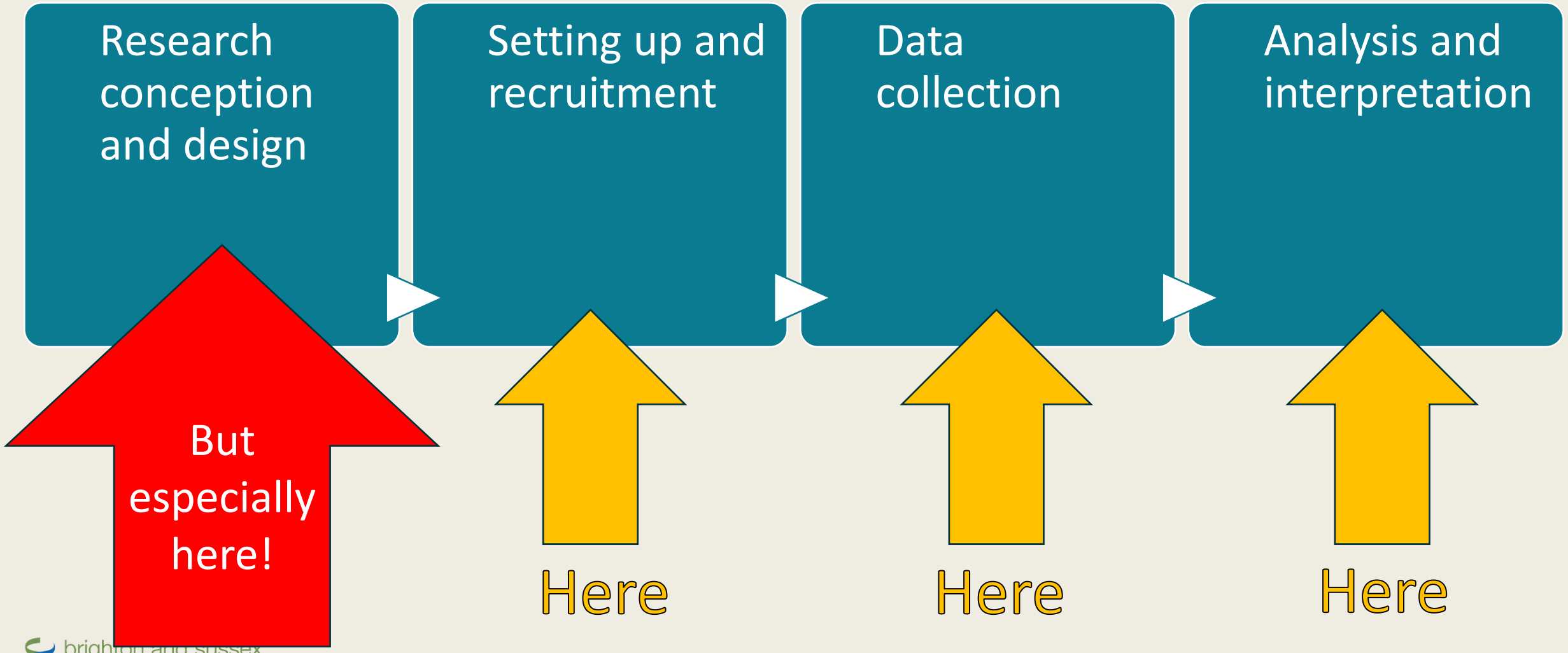
“Dementia is awful. It definitely needs something, and you’ve got to start somewhere, so **I think the assessment is a start**.”

“If the medications work better with an early diagnosis, that’s more power to the assessments.”

What should be the next steps for the dementia prediction tool?

Brainstorm ideas in the chat…

When should we do stakeholder/end-user engagement?



brighton and sussex medical school

Where can I get help if I want to start a health data project?

www.bsms.ac.uk/arcdatahub

Welcome to the ARC KSS Data Science Hub!

An open access resource, identifying and exploring national and regional (Kent, Surrey and Sussex) health and social care datasets. A space where data access barriers are addressed, in the hope of encouraging improved healthcare based on the real needs of everyday people as users of health and care services.

https://www.sussex.ics.nhs.uk/our-work/our-priorities/digital/kent-medway-and-sussex-secure-data-environment-for-research/

Kent, Medway & Sussex
SECURE DATA ENVIRONMENT

brighton and sussex medical school

Thank you – Questions?

Contact me:

e.m.ford@bsms.ac.uk

www.bsms.ac.uk/dr-elizabeth-ford

www.bsms.ac.uk/arcdatahub

NIHR | Applied Research Collaboration Kent, Surrey and Sussex

brighton and sussex medical school