Following the health of half a million participants

Cathie Sudlow

UK Biobank Scientific Conference

London, June 2018

Follow-up of participants in very large prospective cohorts

Aim: identify a wide range of incident diseases and other health related outcomes

Active methods requiring participant re-engagement

- face to face reassessment
- postal or web-based surveys
- with electronic capture can acquire data very rapidly
- prone to incomplete coverage & selective loss to follow-up
- miss cases emerging between assessments

Passive methods via linkages to national health records

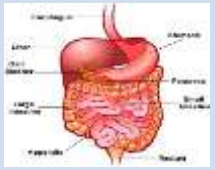
- can follow all participants without need for re-engagement
- should be efficient and cost effective
- broad consent at recruitment important
- rely on universal healthcare system & availability of relevant datasets
- can only detect cases of disease diagnosed in a healthcare setting
- data need to be accurate and sufficiently detailed for research studies

biobank uk

Web questionnaires

- Using email and web questionnaires
  - for more detailed assessment of exposures
  - and to obtain information on outcomes that cannot be obtained through linking to health records
- Of 350,000 with email, 30-50% complete each questionnaire

Web questionnaires

Date	Topic	Numbers (%) completed
2011-2012	24 hour dietary recall (x 4) 	176,012 across 4 rounds (26%-33% for each round)
2014	Cognitive function 	120,852 (36%)
2015	Occupational history 	118,009 (36%)
2016	Mental health 	162,369 (47%)
2017	Digestive health 	172,735 (52%)

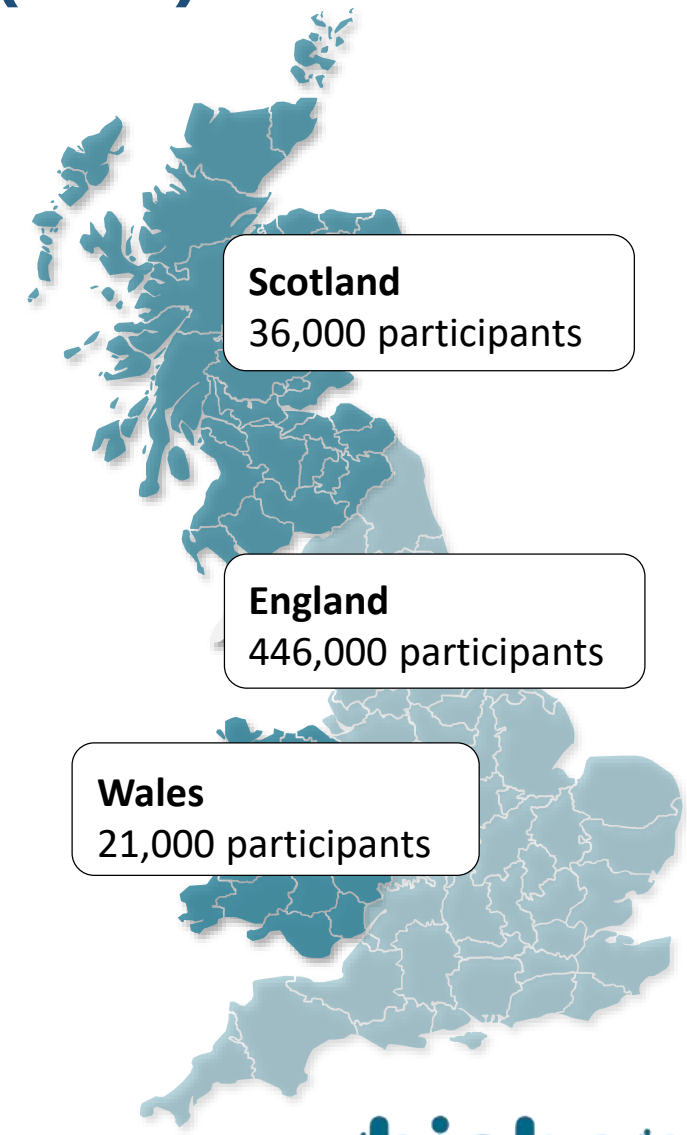
Useful for following change over time…but beware selective attrition

biobank uk

Following the health of 0.5 million UK Biobank participants through linking to National Health Service (NHS) records

Regularly updated information on a wide range of diseases from NHS datasets in all three countries:

- Deaths - date and cause of death
  for all participants
  >14,000 by early 2016
- Cancers – date, stage and grade of cancer
  for all participants
  >79,000 cancer cases by late 2015
- Admissions to hospital – dates, diagnoses, procedures
  for all participants
  1000's of cases of many incident diseases
- Primary care data – dates, diagnoses, symptoms, signs, referrals, prescriptions, labs etc
  for half of the participants
  1000's more cases of many incident diseases



biobank uk

Maximising the value of the linked healthcare data

- Messy ‘real world data’ - not collected primarily for research
- Not 100% accurate due to administrative and clinical error
- Mainly structured, coded datasets (ICD, OPCS4, Read…)
- Experts advising in a range of disease areas:

Cancer
Diabetes
Cardiac diseases
Stroke
Mental health disorders
Eye diseases
Neurodegenerative diseases
Chest diseases
Musculoskeletal conditions
Infections
Kidney diseases

- Combine different linked data sources to create algorithmically derived disease status indicators
- Estimate the accuracy and completeness of these
- Consider limitations and potential additional sources of unstructured data

biobank uk

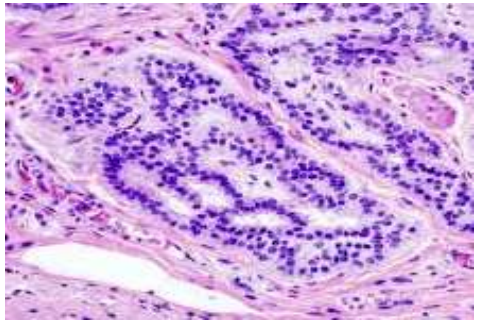
Cancers in UK Biobank ascertained from the national cancer registries

	Observed		Predicted
	By recruitment	Incident by 2015	Incident by 2022
Breast cancer	9,000	4,200	10,000
Colorectal cancer	2,300	2,500	7,000
Prostate cancer	3,000	4,300	9,000

→ Date, stage and grade of cancer

Beyond the structured registry data…obtaining additional information for subtyping of identified cancer cases through regional linkages to:

- histopathology reports
- digitised histopathology slides
- tumour specimens



biobank uk

Numbers of non-cancer disease cases in UK Biobank

ascertained from self-report at recruitment, hospital admissions and death registries

Condition	Prevalent cases observed at recruitment 2006-2010	New cases observed by 2016	New cases predicted by 2021	New cases predicted by 2026
Dementia	200	1,800	5,400	18,000
Stroke	7,800	4,600	8,300	18,400
MI	12,000	7,400	12,200	20,500
COPD	9,800	7,600	13,300	23,800
Parkinson’s Disease	1,000	1000	2,000	4,700

Numbers of non-cancer disease cases in UK Biobank

ascertained from self-report at recruitment, hospital admissions and death registries

ascertained from self-report at recruitment, hospital admissions, death registries and primary care data

Condition	Prevalent cases observed at recruitment 2006-2010		New cases observed by 2016		New cases predicted by 2021		New cases predicted by 2026	
Dementia	200	200	1,800	4,300	5,400	13,000	18,000	43,400
Stroke	7,800	8,400	4,600	7,100	8,300	12,900	18,400	28,500
MI	12,000	12,200	7,400	8,000	12,200	13,300	20,500	22,300
COPD	9,800	11,500	7,600	10,000	13,300	17,500	23,800	31,300
Parkinson's Disease	1,000	1,000	1000	2,000	2,000	4,000	4,700	9,700

How accurate? How complete?

Numbers of non-cancer disease cases in UK Biobank

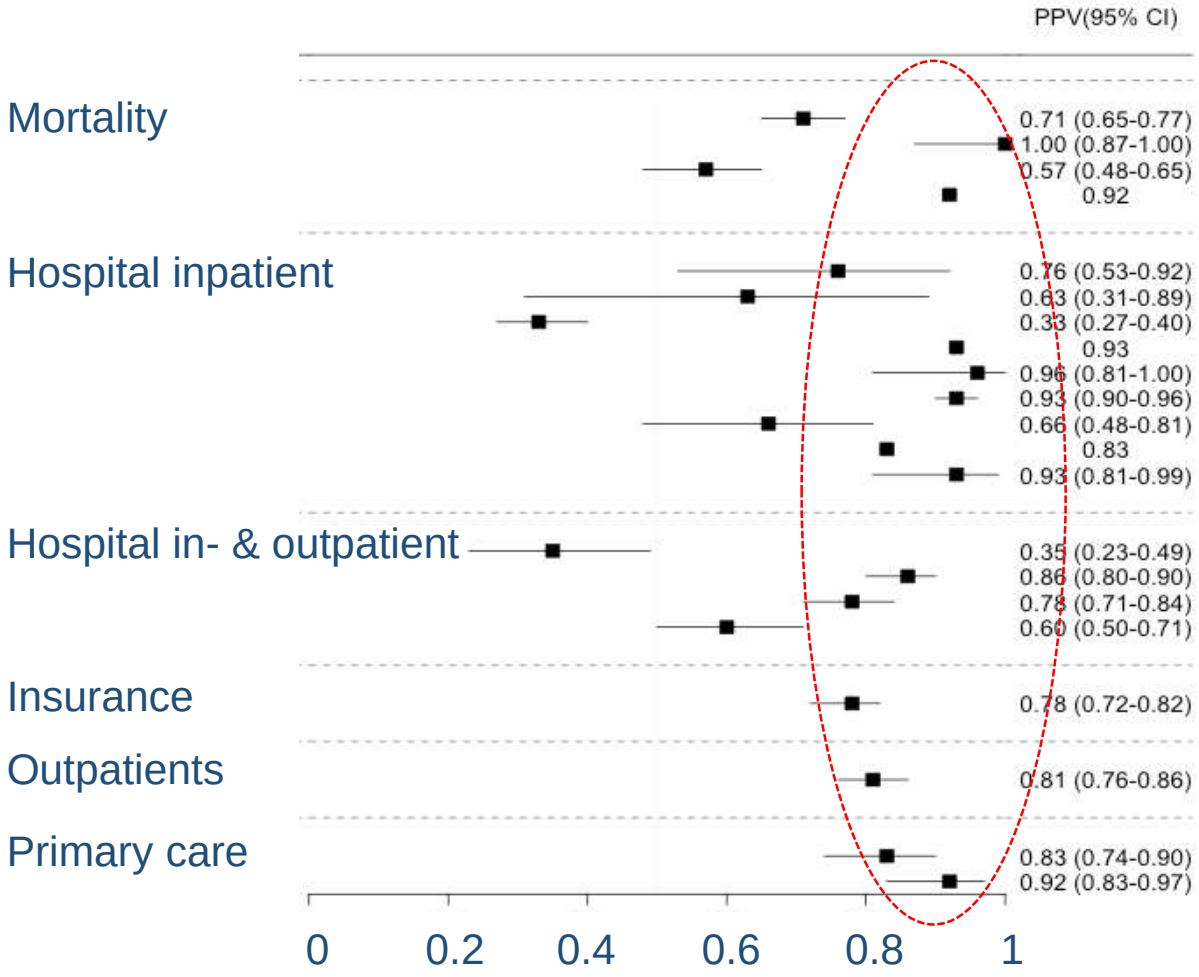
ascertained from self-report at recruitment, hospital admissions and death registries

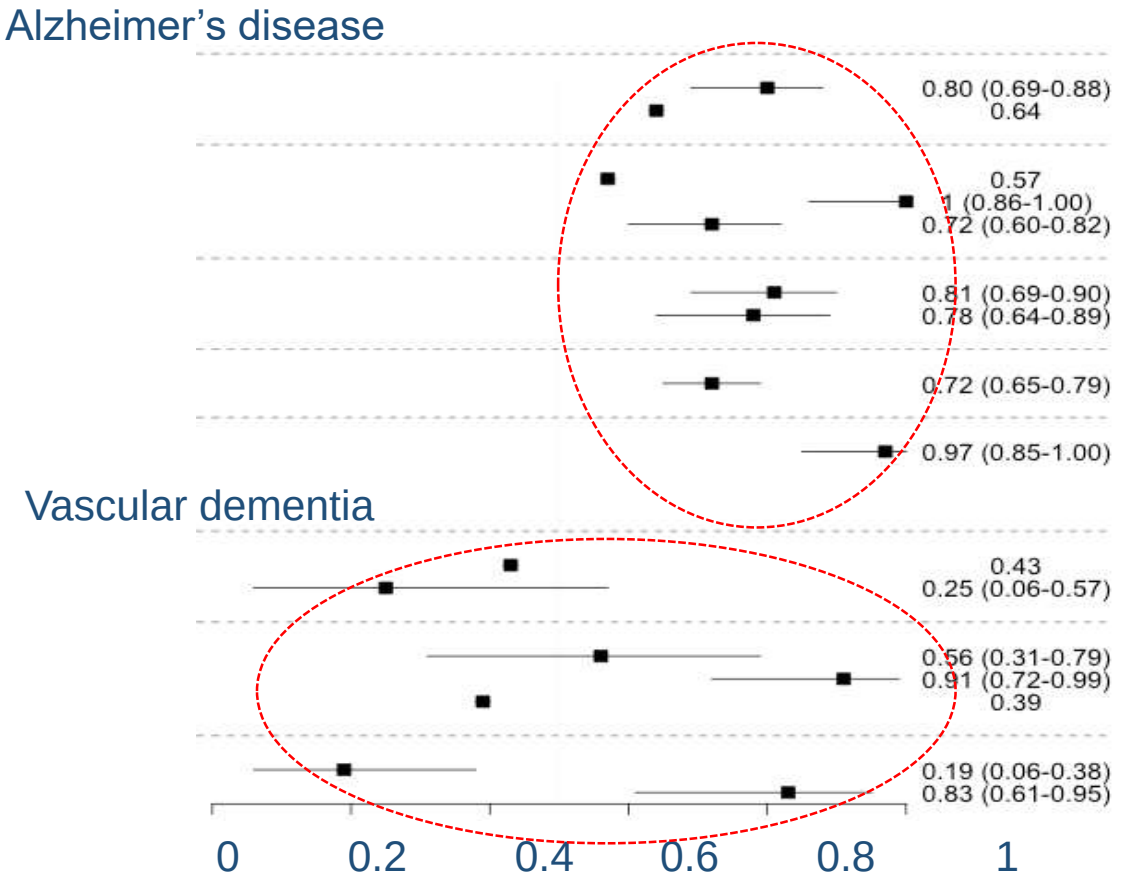
ascertained from self-report at recruitment, hospital admissions, death registries and primary care data

Condition	Prevalent cases observed at recruitment 2006-2010		New cases observed by 2016		New cases predicted by 2021		New cases predicted by 2026	
Dementia	200	200	1,800	4,300	5,400	13,000	18,000	43,400
Stroke	7,800	8,400	4,600	7,100	8,300	12,900	18,400	28,500
MI	12,000	12,200	7,400	8,000	12,200	13,300	20,500	22,300
COPD	9,800	11,500	7,600	10,000	13,300	17,500	23,800	31,300
Parkinson's Disease	1,000	1,000	1000	2,000	2,000	4,000	4,700	9,700

Dementia: positive predictive value of routine healthcare data

From published studies



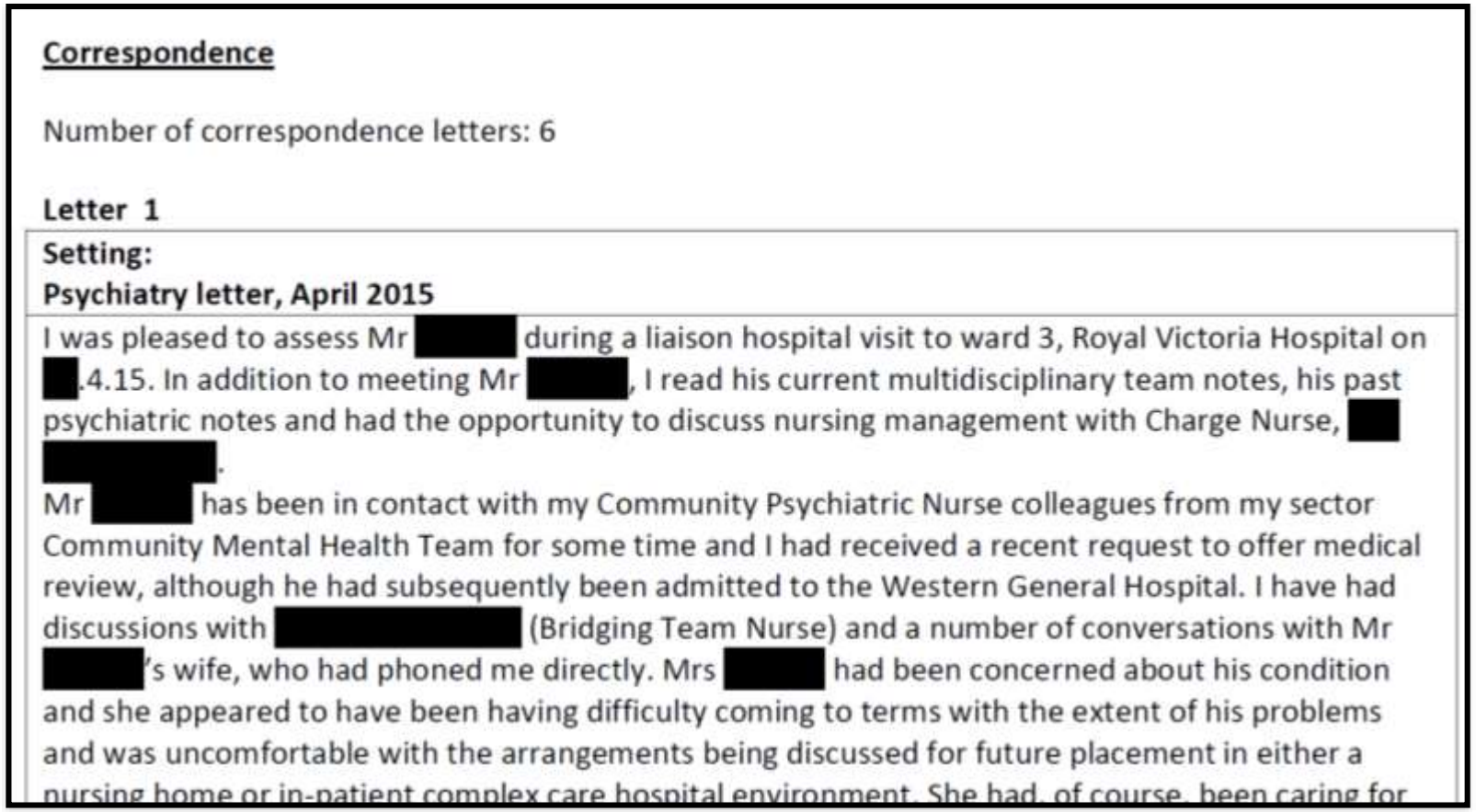


Wide variation but in most PPV >80%

PPV for AD generally higher than for vasc dementia

Dementia: positive predictive value of routine healthcare data

From comparison with expert review of free text electronic medical record



Correspondence

Number of correspondence letters: 6

Letter 1

Setting:
Psychiatry letter, April 2015

I was pleased to assess Mr [redacted] during a liaison hospital visit to ward 3, Royal Victoria Hospital on [redacted].4.15. In addition to meeting Mr [redacted], I read his current multidisciplinary team notes, his past psychiatric notes and had the opportunity to discuss nursing management with Charge Nurse, [redacted].

Mr [redacted] has been in contact with my Community Psychiatric Nurse colleagues from my sector Community Mental Health Team for some time and I had received a recent request to offer medical review, although he had subsequently been admitted to the Western General Hospital. I have had discussions with [redacted] (Bridging Team Nurse) and a number of conversations with Mr [redacted]'s wife, who had phoned me directly. Mrs [redacted] had been concerned about his condition and she appeared to have been having difficulty coming to terms with the extent of his problems and was uncomfortable with the arrangements being discussed for future placement in either a nursing home or in-patient complex care hospital environment. She had, of course, been caring for

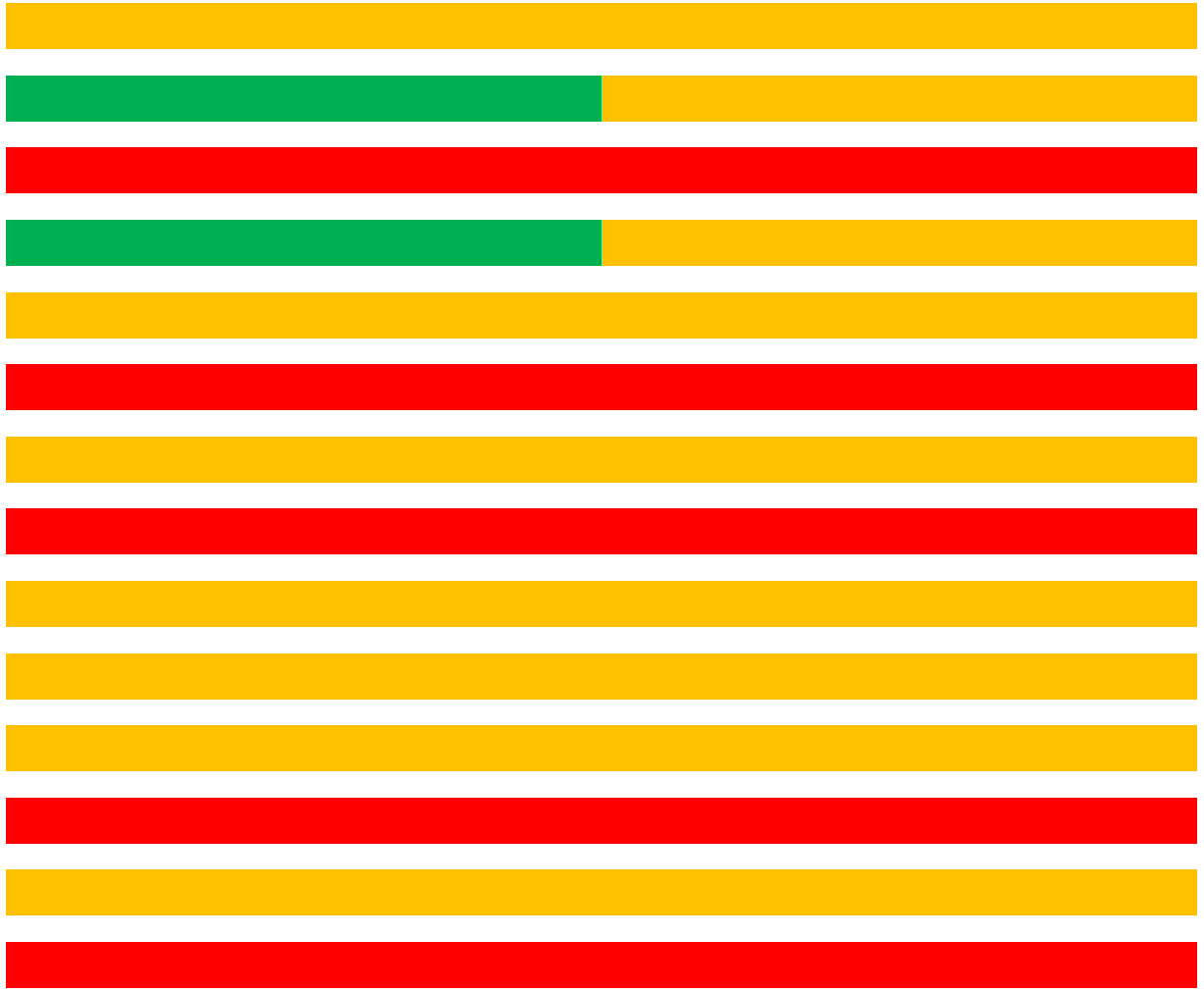
130 cases

Dementia:	80%
Alzheimer's disease:	72%
Vascular dementia:	44%

biobank uk

Non-cancer disease status algorithms: current

Condition	Disease indicator	
	Excluding GP data	Including GP data
Diabetes		
Myocardial infarction		
Other heart conditions		
Stroke		
Fractures		
Arthritis (rheumatoid and osteo-)		
Chronic obstructive pulmonary disease & asthma		
Venous thrombo-embolism		
Inflammatory lung diseases		
Dementias		
Parkinson’s and motor neurone diseases		
Mental health disorders		
End stage renal disease		
Other renal diseases		



Non-cancer disease status algorithms: by end 2018

Condition	Disease indicator	
	Excluding GP data	Including GP data
Diabetes	✓	✓
Myocardial infarction	✓	✓
Other heart conditions		✓
Stroke	✓	✓
Fractures	✓	✓
Arthritis (rheumatoid and osteo-)		✓
Chronic obstructive pulmonary disease & asthma	✓	✓
Venous thrombo-embolism		✓
Inflammatory lung diseases	✓	✓
Dementias	✓	✓
Parkinson’s and motor neurone diseases	✓	✓
Mental health disorders		✓
End stage renal disease	✓	✓
Other renal diseases		✓

Non-cancer disease status algorithms: by end 2019

Condition	Disease indicator	
	Excluding GP data	Including GP data
Diabetes	■	
Myocardial infarction	■	
Other heart conditions		■
Stroke	■	
Fractures	■	
Arthritis (rheumatoid and osteo-)		■
Chronic obstructive pulmonary disease & asthma	■	
Venous thrombo-embolism	■	
Inflammatory lung diseases	■	
Dementias	■	
Parkinson’s and motor neurone diseases	■	
Mental health disorders	■	
End stage renal disease	■	
Other renal diseases		■

Research uses of the linked healthcare data

- Wide range of study types:
  - cross sectional associations
  - large scale genetic analyses
  - prospective epidemiological analyses
- Extent of use depends on:
  - how long the data have been available to researchers
  - pre-existing familiarity with the dataset and coding system
  - complexity of the dataset provided
  - numbers of disease cases and length of follow-up
- Use of prospectively ascertained deaths, cancers and hospital admissions is increasing
- This is particularly the case for the off the shelf cardiovascular outcomes

Beyond the linked coded healthcare data?

Structured, coded data from linked national healthcare datasets:

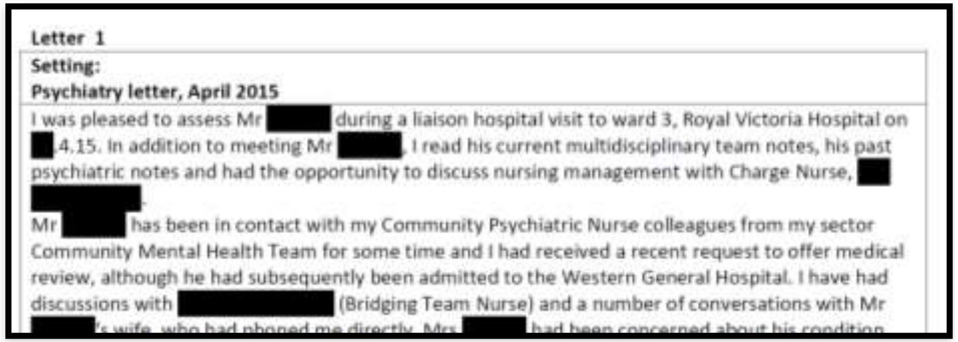
- Can ascertain cases of a wide range of diseases with acceptable accuracy
- Capture only 10-20% of the information from electronic medical records
- Are limited for detailed sub-phenotyping of disease

Deeper phenotyping of disease may require multiple unstructured data sources, including:

- Free text of electronic records
- Complex electrical signalling data (ECG's, EEG's etc)
- Histopathology slide sets
- Clinical imaging data

Obtaining these data at national scale is challenging

To extract value from these data on 1000's of outcomes across multiple diseases, we need scalable approaches: crowd sourcing, natural language processing, machine learning, artificial intelligence…

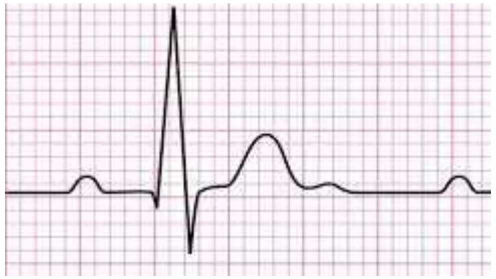

Letter 1

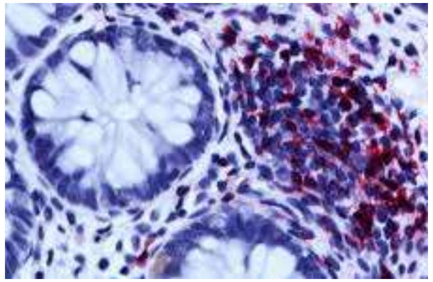
Setting:

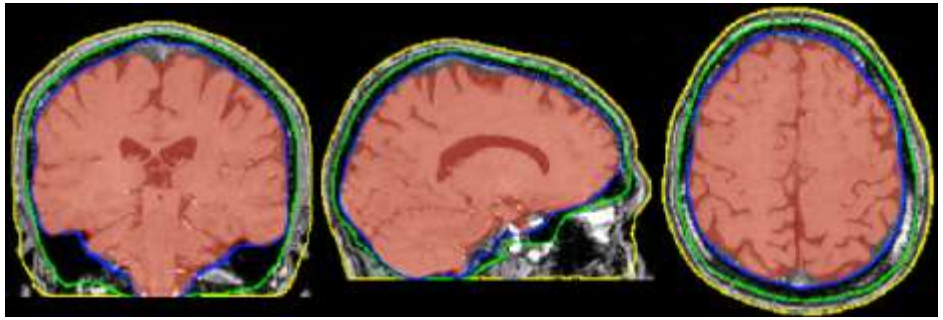
Psychiatry letter, April 2015

I was pleased to assess Mr [redacted] during a liaison hospital visit to ward 3, Royal Victoria Hospital on [redacted]4.15. In addition to meeting Mr [redacted], I read his current multidisciplinary team notes, his past psychiatric notes and had the opportunity to discuss nursing management with Charge Nurse, [redacted]

Mr [redacted] has been in contact with my Community Psychiatric Nurse colleagues from my sector Community Mental Health Team for some time and I had received a recent request to offer medical review, although he had subsequently been admitted to the Western General Hospital. I have had discussions with [redacted] (Bridging Team Nurse) and a number of conversations with Mr [redacted]'s wife, who had phoned me directly. Mrs [redacted] had been concerned about his condition







If you would like to help or make suggestions

Data linkages and algorithms:	cathie.sudlow@ed.ac.uk
Web questionnaires:	naomi.allen@ndph.ox.ac.uk