Mental health in UKBiobank

Professor Matthew Hotopf

Director, NIHR Biomedical Research Centre at the Maudsley

UK Biobank

- 500,000 UK adults age 40-69 at recruitment
- Baseline data on lifestyle, environment, personal & family medical history, physical measures & biological samples
- Follow-up for disease outcomes over 20+ years
- Establish genetic and environmental determinants of common diseases of middle and old age
- Improve prevention, diagnosis and treatment of cancer, heart disease, stroke, arthritis, dementia….

Why is UK Biobank special?

- Very large – will generate large numbers (1000s) of disease cases for adequately powered case-control and case-cohort studies
- Extensive and detailed exposure measures
- Comprehensive follow-up with careful phenotyping of outcomes
- Open access resource: see www.ukbiobank.ac.uk

Recruitment

Participant characteristics

- 46% male
- 57% aged 40-59; 43% aged 60-69
- Less socioeconomically deprived than UK average but all strata represented
- 85% urban
- 94.5% white; 5.5% other
- 58% paid employment / self employed
- 89% recruited in England; 7% in Scotland; 4% in Wales

Automated -80°C sample archive

- Blood
  - whole blood
  - serum
  - plasma
  - red cells
  - buffy coat
- Urine
- Saliva

Total > 15 million aliquots

Other ongoing enhancements

- Web-based questionnaires for additional exposures and outcomes (cognition, mental health, occupation..)
- Repeat of baseline assessment every few years in a subsample of 20-25,000
- Wrist-worn accelerometers mailed to 100,000 participants to measure physical activity
- Standard assay panel on samples from all participants
- Multimodal imaging in 100,000 - brain, cardiac and abdo MRI & 3D carotid ultrasound
- Genotyping of all 500,000 (800+ SNPs)
- Cardiac rhythm monitoring

Long term follow-up in UK Biobank

Value of resource depends on:

- Rich baseline data, samples and further planned enhancements
- Comprehensive and detailed follow-up of health of participants

Key advantages for UK Biobank:

- NHS provides majority of healthcare in the UK
- Cohort-wide linkage to a wide range of routine coded health records possible

Prevalent conditions* at recruitment

Condition	Cases (n)
Diabetes	26,000
MI	12,000
COPD	12,000
Stroke	7,000
Breast cancer	11,000
Colorectal cancer	3,000
Prostate cancer	3,000
Rheumatoid arthritis	6,000
Retinal detachment	2,000

Incident outcomes* during follow-up

Condition	2012	2017	2022
Diabetes	10,000	25,000	40,000
MI/CHD death	7,000	17,000	28,000
Stroke	2,000	5,000	9,000
COPD	3,000	8,000	14,000
Breast cancer	2,500	6,000	10,000
Colorectal cancer	1,500	3,500	7,000
Prostate cancer	1,500	3,500	7,000
Lung cancer	1,000	2,000	4,000
Hip fracture	1,000	2,500	6,000
Alzheimer’s	1,000	3,000	9,000

Record linkages

Complete

- Cancer registration
- Death registration
- Hospital episode statistics

In progress

- GP records (complete in Scotland and Wales)
- Mental health minimum dataset
- IAPT

Strategy for follow up: questionnaires

- Some outcomes of major interest not captured by routine healthcare datasets
- Questionnaires to participants are traditional means of follow-up in many longitudinal studies
- 2/3 of participants have current email address, making email invitations and web administration possible
- Proof of concept: dietary recall questionnaire
- Series of further questionnaires planned:
  - Cognitive function
  - Occupational history
  - Mental health
  - Others: quality of life, disability, site-specific pain, early neurodegenerative symptoms, over-the-counter medications, social interactions and major life events

Adjudication of health outcomes

What do the coded data actually tell us?

How accurate?

How detailed?

How complete?

Do we need to go beyond the coded data?

Expert-led adjudication of outcomes

Outcomes working group

- Developing methods for ascertainment, confirmation, and sub-classification of disease outcomes:

Cancer
Diabetes
Cardiac
Stroke
Mental health
Ocular
Neurodegenerative
Respiratory
Musculoskeletal
Hepatobiliary
Renal

Proposed general principles for outcomes adjudication

Avoid false positive cases
(but tolerate some false negatives)

Staged approaches

Geographical generalisability

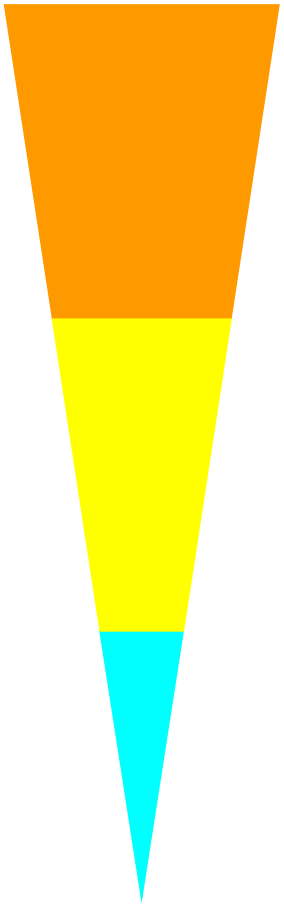
Scalability

Centralised IT development

Cost-effectiveness

Future proof

Staged approaches to outcomes adjudication



APPROACH	CHARACTERISTICS	POSSIBLE EXAMPLES
Ascertainment of suspected cases	Cost-effective Feasible Geographically generalisable Scalable	Cause-specific mortality Cancer incidence Hospital discharge records GP records Web self-report q'aire
Confirmation of "caseness"	As above but somewhat higher cost / lower feasibility	Disease registers Cross-referencing e-records Targeted blood sampling with cheap assays
Sub-classification of cases	More involved and costly	Targeted blood sampling with costly assays Tumour collection / assays Specialised databases (eg imaging) Review of clinical records

Baseline questionnaire

Touchscreen interview

- Socio-demographics and occupation;
- lifestyle exposures (including smoking, alcohol, physical activity and diet);
- early life exposures;
- psychological state;
- cognitive function;
- family history of illness;
- medical history and general health.

Psychological state questions

- Psychological trait (neuroticism) and mood based on standardized questionnaires,
- Serious life events and medical presentations for psychological symptoms.
- Paired associated learning questions to assess global cognition and reaction time tests for touch-screen administration

biobank

Data-Field 20002

Description: Non-cancer illness code, self-reported

Category: Medical conditions - Verbal interview - UK Biobank Assessment Centre

Participants	378,256	Value Type	Categorical (multiple)	Sexed	Both sexes
Item count	990,792	Item Type	Data	Instances	Defined (2)
Stability	Complete	Strata	Derived	Array	Yes (29)

Data | Notes | 4 Categories | 5 Related Data-Fields | 0 Tabulations | 3 Resources

990,792 items of data are available, covering 378,256 participants, encoded using Data-Coding 6.
Defined-instances run from 0 to 1, labelled using Instancing 2.
Array indices run from 0 to 28.

Category	Count
cardiovascular	280034
respiratory/ent	108758
gastrointestinal/abdominal	100923
renal/urology	27158
endocrine/diabetes	60398
neurology/eye/psychiatry	-
neurology	37498
neurological injury/trauma	2880
eye/eyelid problem	25666
psychological/psychiatric problem	491
depression	29726
post-natal depression	452
anxiety/panic attacks	6929
nervous breakdown	732
schizophrenia	636
deliberate self-harm/suicide attempt	222
mania/bipolar disorder/manic depression	1458
substance abuse/dependency	862
post-traumatic stress disorder	363
anorexia/bulimia/other eating disorder	384
stress	715
obsessive compulsive disorder (ocd)	114
insomnia	522
musculoskeletal/trauma	146719
haematology/dermatology	30709
gynaecology/breast	33795
immunological/systemic disorders	48845
infections	17186
unclassifiable	-

Top level
Level 1
Level 2
Level 3
Level 4

Issues for mental health

1. Most disorders have onset in adolescence or early adult life.
2. Disorders fluctuate over time
3. Some disorders inherently difficult to identify in questionnaire surveys
4. Routine diagnosis in linked records may be unreliable
5. Multitude of disorders and exposures
6. Participant fatigue and stigma
7. Low participation rates – differential impact on mental disorders

High participation required to be representative of psychiatric disorder

Diagnosis Group	No. of Participants	No. of Nonparticipants	Association Between Diagnosis and Nonparticipation		% of Cases Who Participated	
			Relative Risk	95% CI	%	95% CI
Current and future DP						
No DP awarded	17,162	9,296	1.00	Referent		
Mental disorders	447	761	2.98	2.87, 3.10	37.0	34.3, 39.7
Musculoskeletal disorders	448	299	1.22	1.08, 1.37	60.0	56.5, 63.5
Other somatic disorders[b]	508	479	1.70	1.58, 1.83	51.5	48.4, 54.6
Current DP[c]						
No DP awarded	17,162	9,296	1.00	Referent		
Mental disorders	279	493	3.15	3.00, 3.29	36.1	32.8, 39.5
Musculoskeletal disorders	143	120	1.54	1.30, 1.78	54.4	48.4, 60.4
Other somatic disorders	253	259	1.87	1.69, 2.04	49.4	45.1, 53.7
Future DP[d,e]						
No DP awarded	17,162	9,296	1.00	Referent		
Mental disorders	168	268	2.89	2.70, 3.08	38.5	34.0, 43.1
Musculoskeletal disorders	305	179	1.08	0.90, 1.26	63.0	58.7, 67.3
Other somatic disorders	255	220	1.58	1.40, 1.76	53.7	49.2, 58.2

Knudsen et al: Am J Epi 2010

Proposed strategy for mental disorders

- Two-pronged approach – combine record linkage with web-based interview
- For psychosis and bipolar – main emphasis on record linkage
- For common disorders – use life time diagnosis from a web-based survey.
- *Emphasise specificity over sensitivity*

Process

- Established a mental health working group by inviting widely.
- Meeting in Jan 2015
- Smaller group to refine questionnaire design
- Iterative process of consultation and refinement
- Ethics approval received April 2016
- Questionnaire piloted and ready to implement

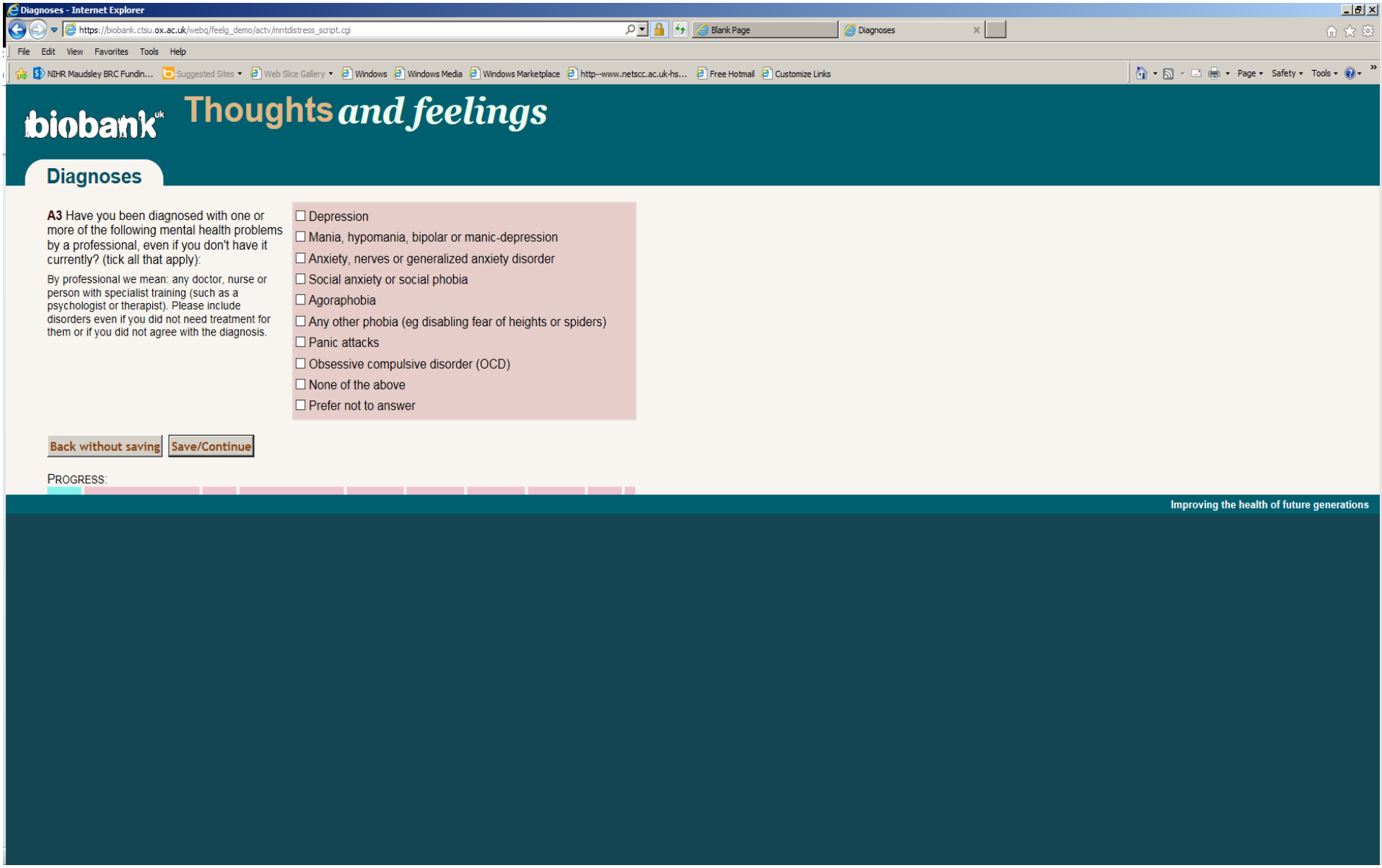
Web-based interview - requirements

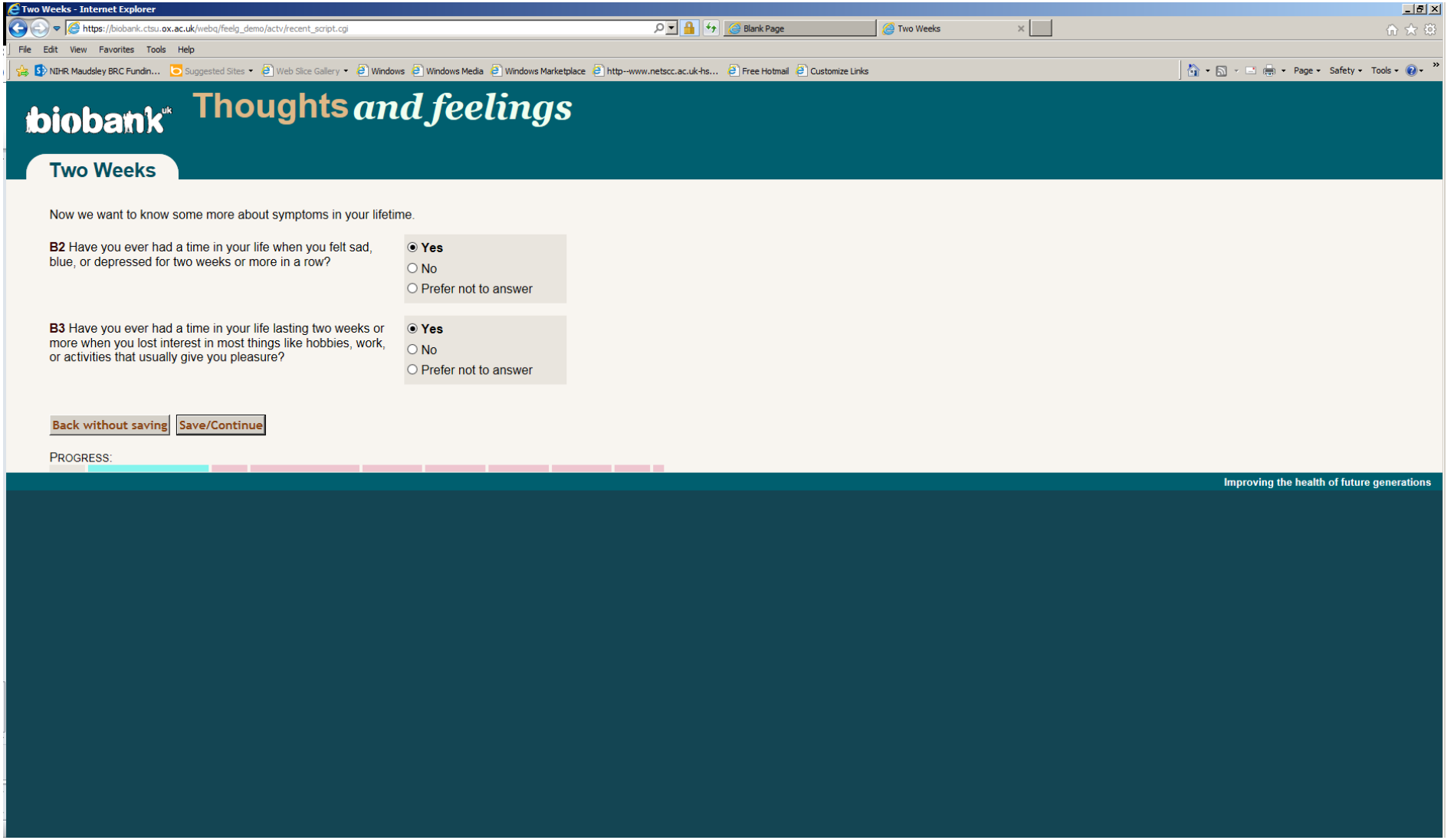
- Quick (20 minutes)
- Acceptable
- Cover main common mental disorders…
- …and important exposures.

- Give *lifetime* as opposed to *current* diagnosis
- Consistent with other major collections: e.g. International Psychiatric Genomics Consortium

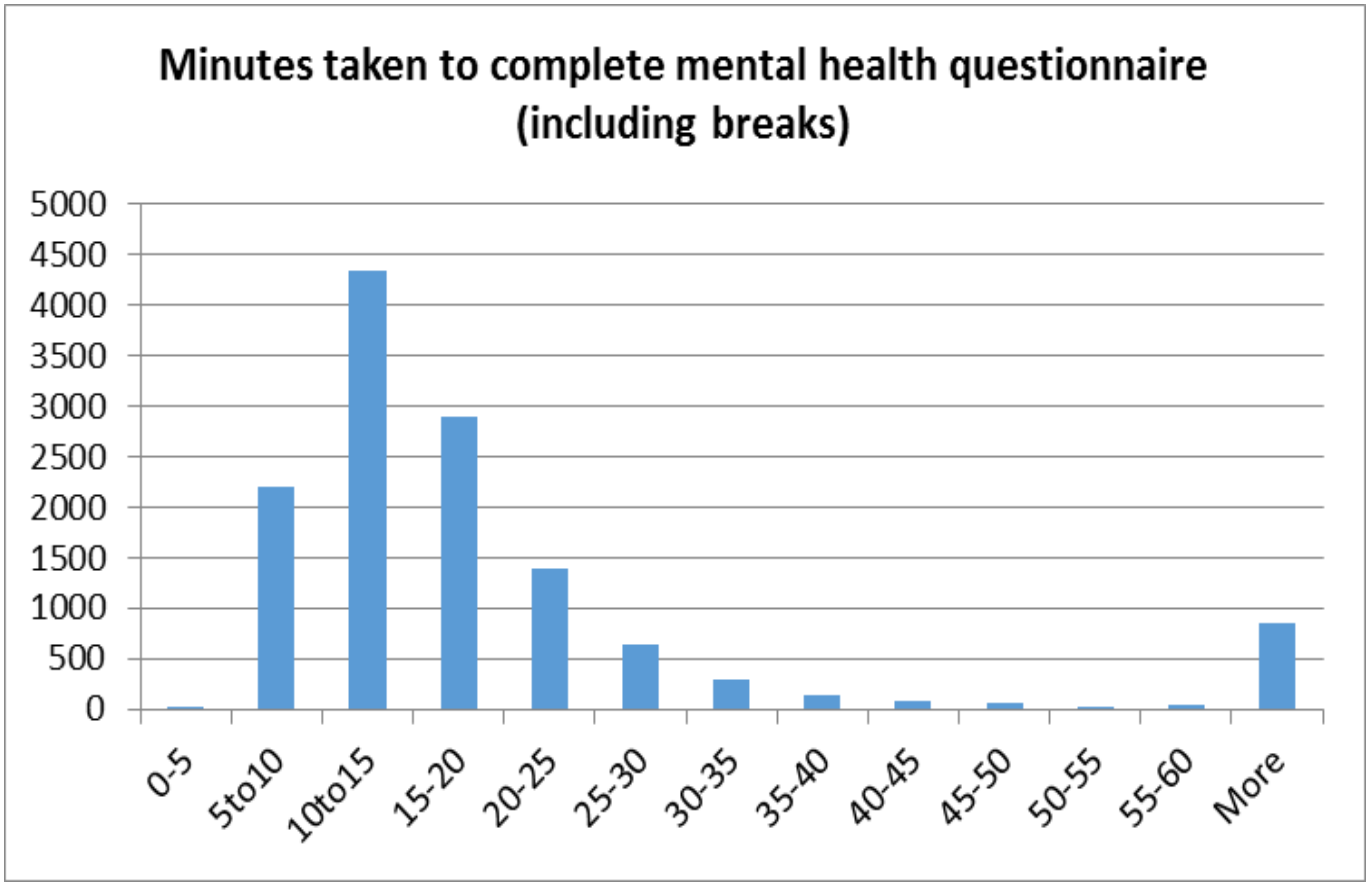
Content

- Lifetime screen question – in your life have you suffered from a period of mental distress that prevented you from doing your usual activities
- Current depression (PHQ-9)
- Lifetime depression (CIDI-SF)
- Brief lifetime mania screen
- Current anxiety (GAD-7)
- Lifetime addictions – alcohol, over-counter, illicit drugs, behavioural
- Current alcohol consumption and use patterns (AUDIT)
- Lifetime cannabis use
- Lifetime psychotic experience screen
- Childhood experiences including abuse
- Lifetime and past 12 month traumatic events
- PTSD symptoms
- Life-time self harm


Diagnoses - Internet Explorer
https://biobank.ctsu.ox.ac.uk/webq/feelg_demo/actv/mntdistress_script.cgi
Blank Page
Diagnoses
File Edit View Favorites Tools Help
NIHR Maudsley BRC Fundin...
Suggested Sites
Web Slice Gallery
Windows
Windows Media
Windows Marketplace
http--www.netscc.ac.uk-hs...
Free Hotmail
Customize Links
Page
Safety
Tools
biobank uk
Thoughts and feelings
Diagnoses
A3 Have you been diagnosed with one or more of the following mental health problems by a professional, even if you don't have it currently? (tick all that apply):
By professional we mean: any doctor, nurse or person with specialist training (such as a psychologist or therapist). Please include disorders even if you did not need treatment for them or if you did not agree with the diagnosis.
Depression
Mania, hypomania, bipolar or manic-depression
Anxiety, nerves or generalized anxiety disorder
Social anxiety or social phobia
Agoraphobia
Any other phobia (eg disabling fear of heights or spiders)
Panic attacks
Obsessive compulsive disorder (OCD)
None of the above
Prefer not to answer
Back without saving
Save/Continue
PROGRESS:
Improving the health of future generations


Two Weeks - Internet Explorer
https://biobank.ctsu.ox.ac.uk/webq/feelg_demo/actv/recent_script.cgi
Blank Page
Two Weeks
File Edit View Favorites Tools Help
NIHR Maudsley BRC Fundin...
Suggested Sites
Web Slice Gallery
Windows
Windows Media
Windows Marketplace
http--www.netscc.ac.uk-hs...
Free Hotmail
Customize Links
Page
Safety
Tools
biobank uk
Thoughts and feelings
Two Weeks
Now we want to know some more about symptoms in your lifetime.
B2 Have you ever had a time in your life when you felt sad, blue, or depressed for two weeks or more in a row?
Yes
No
Prefer not to answer
B3 Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?
Yes
No
Prefer not to answer
Back without saving
Save/Continue
PROGRESS:
Improving the health of future generations

PROTECT study data



Assuming those who took >1hr were having a break:

- Median time to complete 15 minutes
- 75% in under 21 minutes
- 14836 opened 14656 completed (1.2% drop out)

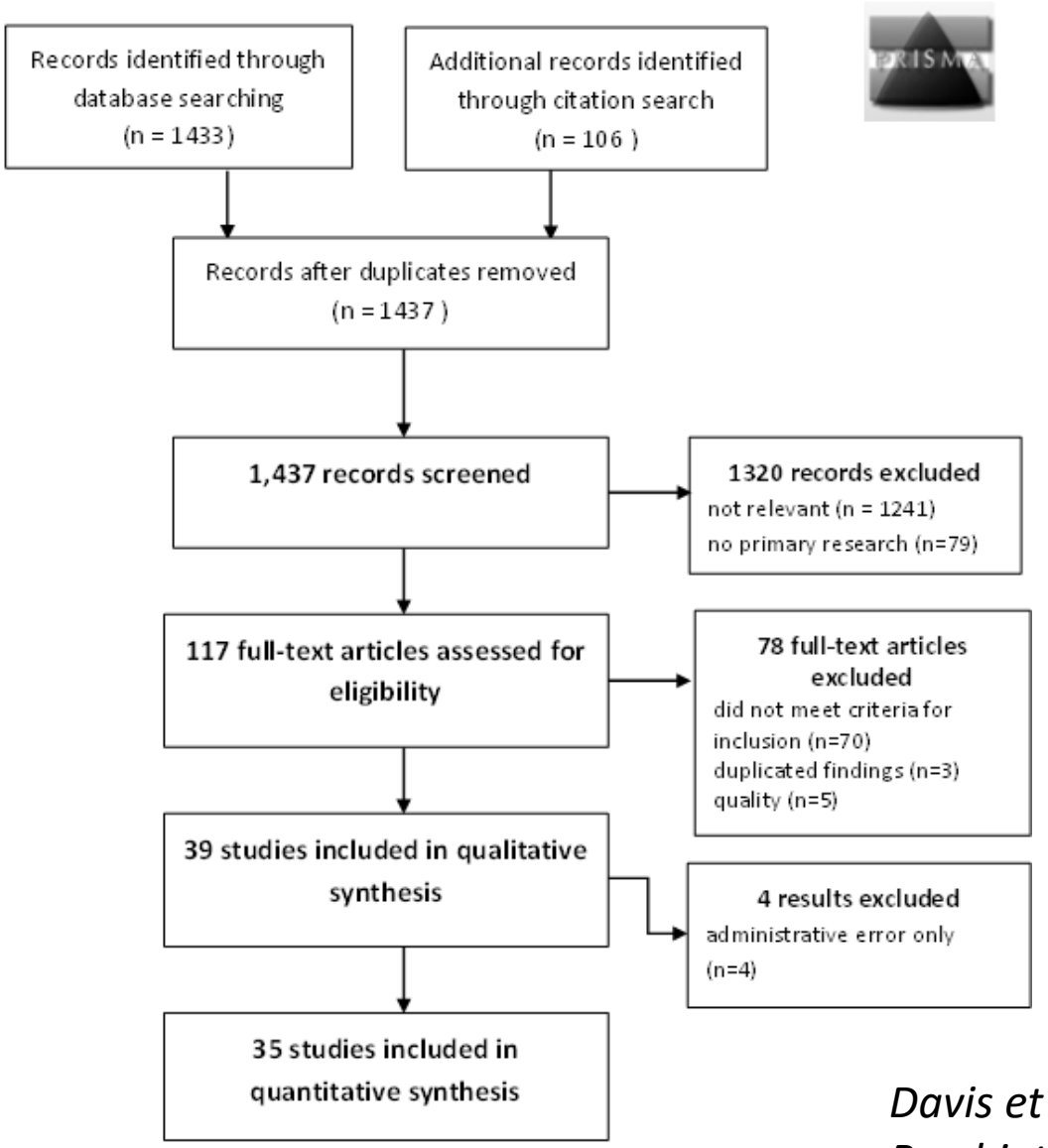
Start of a good weekend…

Hi Matthew,
Quick update:
324794 Emails sent
136218 Viewed
132910 Started
128967 Completed
39.71% Response rate
So likely to hit 40% soon.
Have a good weekend,
Jo

Hospital episode statistics

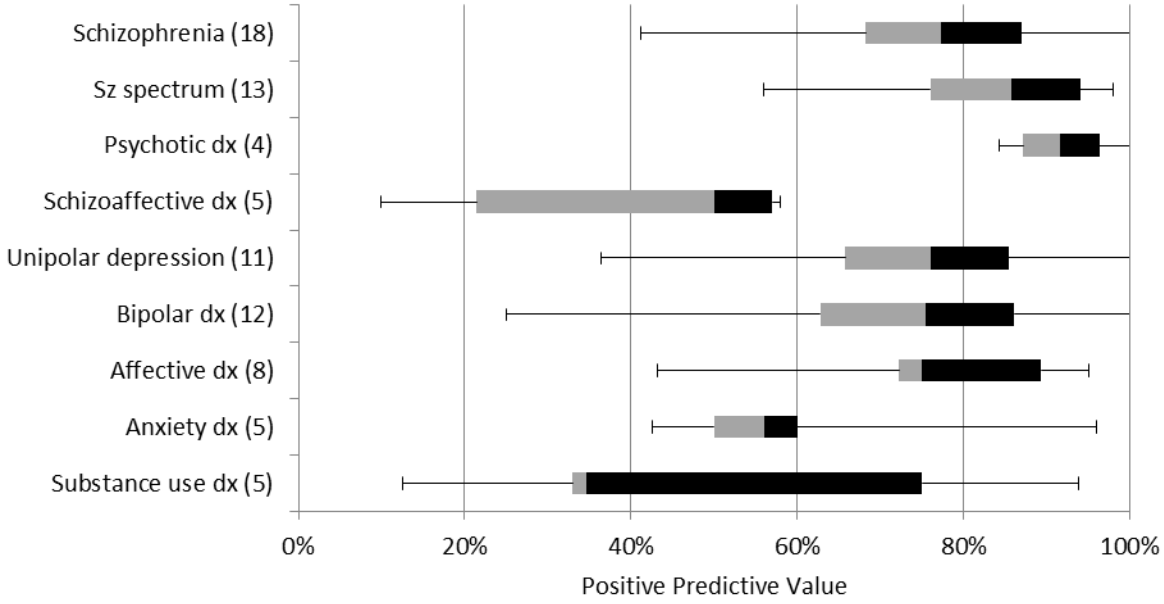
Health outcome	ICD-10 code(s)	Pre-recruitment		Post-recruitment		Total Participants
		Any diagnosis field	Main diagnosis field only	Any diagnosis field	Main diagnosis field only	
Diabetes mellitus	E10-E14, O24, R73.0	10,950	1,600	4,600	350	15,550
Ischaemic heart disease	I20-I25	19,630	15,720	5,460	3,910	25,100
- Acute MI	I21	4,890	4,610	1,340	1,190	6,230
Mood disorders	F30-39	5,050	1,530	2,440	200	7,490
- Unipolar depression	F32-F33	4,530	1,100	2,320	150	6,850
Neurotic disorders	F40-48	2,410	900	1,370	180	3,780
Dementia & Alzheimer's disease	F00-F03, G30	80	20	170	40	240

Validity of administrative data



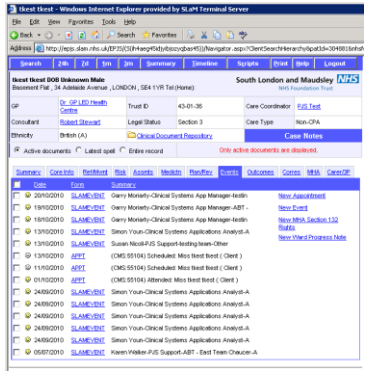
Davis et al, in press BMC Psychiatry

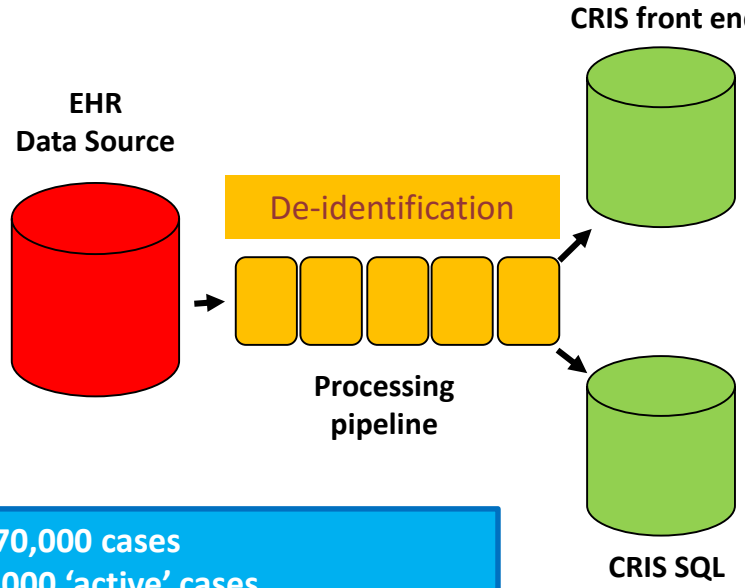
Validity of administrative data

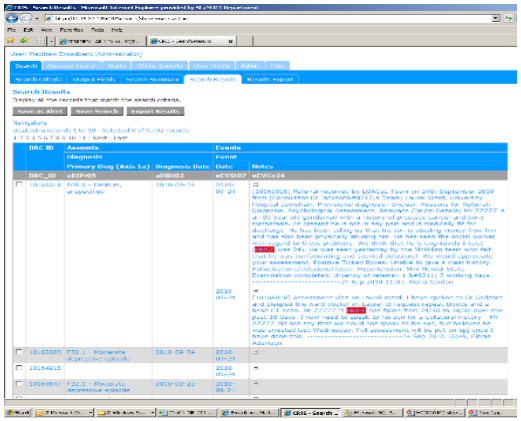


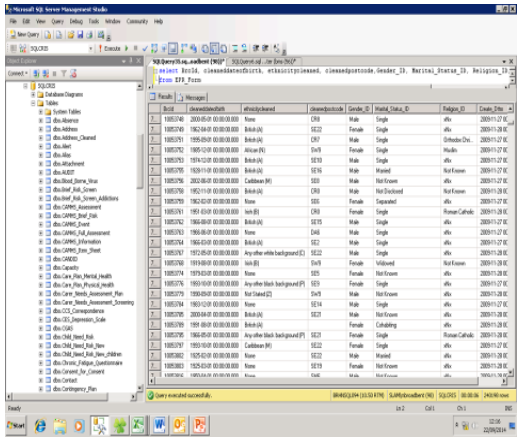
Davis et al, in press BMC Psychiatry

Electronic Patient Records: CRIS – core functionality









~270,000 cases
35,000 'active' cases
125 tables
6500 fields

CRIS - results

	Dementia (n=242)	Schizophrenia (n=1902)	Bipolar (n=910)	Uni Depress (n=952)	Overall (n=4010)
Comparing HES to subsequent HES	82	1220	574	449	4010
Match	94% (86-98)	90% (88-92)	87% (84-90)	72%* (67-76)	89% (87-90)
Comparing HES to structured diagnosis on CRIS	239	1766	829	901	3739
PPV	91%* (86-94)	83%* (81-85)	78% (75-81)	68%* (64-71)	79% (78-80)
Comparing HES to full-text diagnosis on CRIS	212	1520	683	642	3061
PPV	88%* (83-92)	84%* (82-85)	62%* (58-66)	64%* (60-68)	76% (74-77)
Comparing HES to [structured or full text]	239	1766	829	901	3739
PPV	95% (91-97)	93%* (92-94)	89% (87-91)	83%* (80-85)	90% (89-91)
Most common mismatch disorders (≥5% in any comparison)		Other F2 dx	Schizo-phrenia Other F2 dx	Personality Substance use dx	Other F2 dx Substance use dx

Challenges

- Integration of multiple data sources
- Handling missing phenotypes
- Gaining better environmental measures