Mental health phenotyping in UK Biobank

Katrina Davis MB BChir, MRCPsych, Matthew Hotopf CBE, PhD, FRCPsych, FMedSci

Between 2006 and 2010, over half a million people from cities across the UK signed up to be a part of a project to ‘improve the health of future generations’. These participants became the subjects of UK Biobank (UKB), which is now a well-characterised cohort that any scientist can study for health-related research, at a fraction of the cost of setting up a specific study.1-3

UKB was designed to study common diseases of middle and old age, where large sample sizes are needed to study the complex relationship of multiple small-effect genetic and environmental influences. Participants visited an assessment centre where they had baseline measures and gave biosamples, with follow-up for the cohort by data linkage to routine health care data, allowing for efficient and near-complete coverage. Many have participated further, for example through online questionnaires and activity monitoring.

As the model has been proven and technology moved on, ambitions have grown at UKB. For example, initially genotyping was carried out on part of the cohort, but full cohort genotyping became available last year, and now whole genome sequencing is underway. UKB is also leading the way with large-scale multimodal imaging data, captured and processed using homogenous processes.4

UK Biobank and mental health

UKB is a powerful resource for looking at disorders of complex and polygenic aetiology and connections between disorders. Mental disorders and physical-mental comorbidity obviously fit this description. However, there are multiple challenges in characterising mental disorders in cohorts such as UKB.5-7 Firstly, gold-standard clinical interviews would be prohibitively expensive at this scale; secondly, mental health is probably under-recognised and under-reported in routine health care data; thirdly, mental disorders tend to start early in life and fluctuate, such that ascertaining status at one point in time might miss lifetime disorders. The UKB baseline assessment touched upon lifetime mental health with the Eysenck neuroticism scale and a question on help-seeking,8,9 and additional questions in the latter stages of people being enrolled allowed classification of probable lifetime affective disorders in 122 983 people.10 However, given the importance of mental health11-15 it was felt this could be vastly improved; and the UK Biobank Mental Health Outcomes Consortium (UKB-MHO) was formed to help with this. The UKB-MHO is formed of representatives from the UK mental health research community with scientists from UKB, and aims maximise the usefulness of mental health outcomes for all UKB researchers.

Validation and adjudication of routine data

UKB has a virtual registry of hospital admissions, which is the combined data from linkages to Hospital Episode Statistics (HES) for England, Scottish Morbidity Record (SMR) and Patient Episode Database for Wales (PEDW)14 with ICD-10 codes to categorise the main reason for each admission and secondary diagnoses. To help inform the interpretation of these diagnoses for mental health disorder, members of the UKB-MHO carried out a systematic review of the validity of administrative data for diagnosis of mental disorders,15 and subsequently, a set of HES diagnoses was validated against clinical notes.16 This demonstrated that codes indicating an admission to a mental health ward with a diagnosis of schizophrenia, a wider schizophrenia spectrum disorder, bipolar affective disorder or unipolar depression had a positive predictive value of 73% for having a disorder in the same category, but that this varied by diagnosis, schizophrenia being the most reliable category with a positive predictive value of 90%.

Admission due to a mental health disorder will be quite an insensitive measure of mental disorder, and we look forward to a UKB primary care linkage (scheduled to take place during 2019). Members of UKB-MHO also took part in another external validation exercise for GP Read codes relating to depression and anxiety against survey data, finding a set of codes with a positive predictive value of 74% with sensitivity of 32%.17 UKB-MHO is also helping to create algorithms that will combine multiple data sources into adjudicated mental disorder categories, which will facilitate non-specialists in using mental disorders in their research.

Questionnaire development

While participants with mental disorders of high severity can be extracted through linkage, another approach is needed for the majority of people with mental disorder: hence a mental health questionnaire (MHQ) on symptoms of mental disorder was proposed. The questionnaire had to balance the need for depth and breadth with likely
participant preferences (time to complete, ease of use, etc). The characterisation of depression was prioritised, as it was likely to represent the greatest burden in the cohort, then other common disorders, cross-disorder features, and risk factors not captured at baseline. The Composite International Diagnostic Interview Short Form (CIDI-SF) modified to provide lifetime history had been used to identify cases and controls for some existing studies in the Psychiatric Genomics Consortium, so modules from this were chosen to maximise the scope for international collaboration. A service user advisory group at the National Institute of Health Research Biomedical Research Centre at the Maudsley was consulted over the design of the questionnaire and invitation and it was piloted in nearly 15,000 people over 50 years old through an online platform. The MHQ takes an average (median) of 15 minutes to complete and was approved as a substantial amendment to UKB ethics approval (North West – Haydock Research Ethics Committee, 11/NW/0382). All UKB participants with an email address (n=339,092) were sent a link to the MHQ, and 46% (157,366) had responded by the time of the first analysis (31% of total cohort). This makes it one of the largest mental health surveys ever conducted.

Participants in the MHQ ranged from 42 to 81 years old, with a median age of 65 years. Criteria for a lifetime mental disorder was present in 43% (70,892), most commonly depression (24%, 37,434), with extensive comorbidity between conditions. Those who met criteria for a disorder were more likely to be female (except for alcohol use disorder), have high neuroticism, report adverse life events, have a long-term illness and live in an area of high deprivation – all of which add to the face validity of the questionnaire.

The self-completion of the questionnaire, compared with the depth of phenotyping that could be gained from a gold-standard clinical interview, may be a limitation, particularly when dealing with related, but distinct, phenotypes such as psychotic experience and symptoms of psychosis. But genetic studies and their meta-analyses have shown themselves to be fairly robust to different

<table>
<thead>
<tr>
<th>Area</th>
<th>Instrument</th>
<th>Diagnosis phenotype</th>
<th>Cross-cutting phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood disorders</td>
<td>Patient Health Questionnaire – (PHQ-9)</td>
<td>Lifetime depression</td>
<td>Wider bipolar symptoms</td>
</tr>
<tr>
<td></td>
<td>Lifetime Composite International Diagnostic Interview Short Form (CIDI-SF)</td>
<td>Severe lifetime depression</td>
<td>Treatment</td>
</tr>
<tr>
<td></td>
<td>Major Depressive Disorder</td>
<td>Sub-threshold lifetime depression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mania questions</td>
<td>Current depression</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bipolar affective disorder type 1</td>
<td></td>
</tr>
<tr>
<td>Generalised</td>
<td>Generalised Anxiety Disorder Questionnaire (GAD-7)</td>
<td>Lifetime generalised anxiety disorder</td>
<td>Treatment</td>
</tr>
<tr>
<td>anxiety</td>
<td>Lifetime CIDI-SF Generalised Anxiety Disorder</td>
<td>Current generalised anxiety disorder</td>
<td></td>
</tr>
<tr>
<td>Substance use</td>
<td>Alcohol Use Disorder Identification Tool (AUDIT)</td>
<td>Addiction</td>
<td>Cannabis use</td>
</tr>
<tr>
<td></td>
<td>Cannabis use (intensity and recency)</td>
<td>1y alcohol use disorder</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>severe alcohol use disorder</td>
<td></td>
</tr>
<tr>
<td>Unusual experiences</td>
<td>Lifetime Composite International Diagnostic Interview Psychosis (adapted)</td>
<td></td>
<td>Lifetime psychotic experience</td>
</tr>
<tr>
<td>Life events</td>
<td>Childhood Trauma Screen (CTS)</td>
<td>Current post-traumatic stress disorder</td>
<td>Treatment</td>
</tr>
<tr>
<td></td>
<td>Adulthood Questions</td>
<td></td>
<td>Child abuse and neglect</td>
</tr>
<tr>
<td></td>
<td>Post-Traumatic Stress Disorder CheckList -5 (PCL-5)</td>
<td></td>
<td>Adult domestic violence and hardship</td>
</tr>
<tr>
<td>Self-harm</td>
<td>Self-harm</td>
<td></td>
<td>Lifetime and 12m non-suicidal self-injury</td>
</tr>
<tr>
<td>Wellbeing</td>
<td>2 x euphoric, 1 x eudemonic questions</td>
<td></td>
<td>suicidal ideation / suicide attempt</td>
</tr>
</tbody>
</table>
Methods of ascertainment of categorical mental disorder diagnosis,23,24 showing that large samples have a role, even when of relatively poor resolution.

Representativeness
Given that participation in research can be patterned by mental health,25,26 it is essential that researchers in this field consider how the nature of participation may affect their results. In epidemiological surveys there is a choice to be made about whether to strive for representativeness, or gather data from more people at less cost per participant. UK Biobank aimed at scale, and has a selection/participation bias.27 Selective participation in the MHQ then increased this. MHQ participants were on average better educated, of higher social class, less likely to be a current smoker or report a longstanding illness/disability than people of their age in the general population.6 This lack of representativeness means the results cannot be used to estimate prevalence. Measures of association do not rely upon representative samples, but because some characteristics in UK Biobank are highly selected (for example, only 7% of people completing the mental health questionnaire had no qualifications compared with approximately 38% of people in this age range in the UK), there is potential for so-called ‘collider bias’ between variables.28

Opportunities
So far, the greatest interest in UK Biobank has been to look at genetics (839 applications received before July 2018).29 Genotyping was performed with the Affymetrix UK Biobank Axiom array, and imputation leading to potential information on over 90 million single nucleotide polymorphisms (SNPs), indels and large structural variants. In mental health, the Psychiatric Genomic Consortium has played a role in integrating genome-wide association study (GWAS) results into larger evidence base for depression25 and alcohol use disorder,30 and we expect further publications for other disorders.

There has also been much interest in a forthcoming biomarker panel that includes some markers that have been associated with mental disorders, such as C-reactive protein and vitamin D.31 This year also saw the release of tranche of brain imaging data, from a project that aims to eventually cover 100 000 participants.1 This includes three structural modalities, resting and task-based fMRI, and diffusion MRI; followed by an automated image processing pipeline that produces distinct brain structure and function measures, such as grey matter volume in different regions, and connectivity between areas. Other researchers have processed the activity information, available as wrist-worn accelerometer data for 100 000 participants over one week,32 which has great potential to reveal behavioural phenotypes.

All of this work, and that of the UKB-MHO, is laying the groundwork for discoveries into aetiology and pathways to differential outcomes that will influence future preventative and treatment strategies, not only for mental disorders, but the physical disorders that are connected. The self-reported phenotype with biochemical markers, genotype, image-based phenotype and behavioural phenotype combined in the same individuals may be able to approach and address many issues. UKB can provide all of these data, but the evaluation will need computing power, analytical skills, and experts from different fields coming together. To make the most of this large and rich data source will require researchers with the curiosity to ask the questions about mental health that matter, with the imagination and openness to adapt methods to play to the strengths of the data.

Acknowledgement
UK Biobank data for this paper is from application 16577, with thanks to Gerome Breen and Jonathan Coleman for help with accessing and processing data. Thank you to all of the past and present UK Biobank Mental Health Outcomes Consortium. This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Declaration of interest
Matthew Hotopf reports grants for IMI RADAR-CNS and personal fees as an expert witness outside the submitted work.

Dr Davis is Research Associate and NIHR Doctoral Trainee and Professor Hotopf is Professor of General Hospital Psychiatry and Director of the NIHR Maudsley Biomedical Research Centre, both at KCL Institute of Psychiatry, Psychology and Neuroscience, London and South London and Maudsley NHS Trust

References
5. Newson RS, Karlsson H, Tiemeier H. Epidemiological fallacies of