DOI:
10.1002/mpr.1796

Document Version
Peer reviewed version

i)

Indicators of Mental Disorders in UK Biobank – A comparison of approaches

ii)

Indicators of Mental Disorders in UK Biobank

iii)

Katrina A. S. Davis[1,2], Breda Cullen[3], Mark Adams[4], Anamaria Brailean[1], Gerome Breen[1,2], Jonathan R. I. Coleman[1,2], Alexandru Dregan[1], Héléna A. Gaspar[1,2], Christopher Hübel[1,2], William Lee[5,6], Andrew M McIntosh[4], John Nolan[4,7], Robert Pearsall[3] & Matthew Hotopf[1,2]

iv)

1. King's College London, Institute of Psychiatry Psychology and Neuroscience, London, UK

2. South London and Maudsley NHS Foundation Trust, NIHR Biomedical Research Centre, London, UK

3. University of Glasgow, Mental Health and Wellbeing , The Academic Centre, Gartnavel Royal Hospital , Glasgow , Scotland, UK

4. University of Edinburgh, Division of Psychiatry, Edinburgh, UK

5. Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK

6. Devon Partnership NHS Trust , Exeter, UK.

7. UK Biobank

v)

Dr Katrina Davis katrina.davis@kcl.ac.uk

vi) Acknowledgements

This work was supported by the NIHR Biomedical Research Centre at South London and Maudsley and Kings College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the National Institute of Health Research, the Department of Health, or UK Biobank.

High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Other funding:

MA is supported by a Wellcome Trust Strategic Award (Reference 104036/Z/14/Z).

BC is funded by the Scottish Executive Chief Scientist Office (DTF/14/03) and by The Dr Mortimer and Theresa Sackler Foundation.

WL is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula.

AM is supported by a Wellcome Trust Strategic Award (Reference 104036/Z/14/Z) , MRC Grant (Reference MC_PC_17209) and by The Sackler Trust

This work came out of discussions of the UK Biobank Mental Health Outcomes Consortium: Matthew Hotopf (King's College London / South London and Maudsley), Katrina Davis (King's College London / South London and Maudsley), Mark Adams (Edinburgh), Naomi Allen (Oxford / UK Biobank), Gerome Breen (King's College London / South London and Maudsley), Jonathan Coleman (King's College London / South London and Maudsley), Chris Dickins (Exeter), Alexandra Dregan (King's College London), Robin Flaig (Edinburgh / UK Biobank), Elaine Fox (Oxford), Nicholas Graham (Glasgow), Jo Holliday (Oxford / UK Biobank), Louise M Howard (King's College London), Ann John (Swansea), William Lee (Plymouth), Rose McCabe (Exeter), Andrew McIntosh (Edinburgh), Robert Pearsall (Glasgow), Daniel Smith (Glasgow), Cathie Sudlow (Edinburgh / UK Biobank), Joey Ward (Glasgow), Stan Zammit (Bristol / Cardiff).

Declaration of potential conflicts of interest:

We have read and understood the author guidelines of ethical conduct and wish to declare the following:

BC reports grants from the Scottish Executive Chief Scientist Office during the conduct of the study.

MH reports grants for IMI RADAR-CNS and personal fees as an expert witness outside the submitted work.
GB reports grants from National Institute for Health Research during the conduct of the study; support from Illumina Ltd. and the European Commission outside the submitted work.

vii)Abstract

Objectives: For many research cohorts, it is not practical to provide a "gold-standard" mental health diagnosis. It is therefore important for mental health research that potential alternative measures for ascertaining mental disorder status are understood.

Methods: Data from UK Biobank in those participants who had completed the online Mental Health Questionnaire (n=157,363) were used to compare the classification of mental disorder by four methods: symptom-based outcome (self-complete based on diagnostic interviews), self-reported diagnosis, hospital data-linkage and self-report medication.

Results: Participants self-reporting any psychiatric diagnosis had elevated risk of any symptom-based outcome. Cohen's kappa between self-reported diagnosis and symptom-based outcome was 0.46 for depression, 0.28 for bipolar affective disorder, and 0.24 for anxiety. There were small numbers of participants uniquely identified by hospital data-linkage and medication.

Conclusion: Our results confirm that ascertainment of mental disorder diagnosis in large cohorts such as UK Biobank is complex. There may not be one method of classification that is right for all circumstances, but an informed and transparent use of outcome measure(s) to suit each research question will maximise the potential of UK Biobank and other resources for mental health research.

keywords: cohort study, online survey, mental disorder, epidemiology, diagnosis, UK Biobank

viii)

Introduction

Mental health is a major, and growing, contributor to disability worldwide (Whiteford, Ferrari, Degenhardt, Feigin, & Vos, 2015), prompting the need to take advantage of all available resources in order to progress the understanding of mental disorders and the interplay between mental and physical health (Prince et al., 2007). To this end, it is necessary to describe mental disorders and related traits in large-scale epidemiological studies. The use of self-report diagnosis, administrative data and on-line surveys are potential sources of data on mental disorders that may be of use in this context, and so it is important to understand the advantages and limitations of these measures.

Considerations Regarding Indicators of Mental Health

Mental disorder diagnosis is a complex specialist task, requiring elucidation of symptoms, time-course, and context (Casey & Kelly, 2007). It has not yet been possible to categorise mental disorders using pathology or etiology, so, in order that there can be a common language, they have been systematised into syndromes based on signs and symptoms (Clark, Cuthbert, Lewis-Fernández, Narrow, & Reed, 2017). However, it is not clear to what extent these syndromes reflect true disease entities, leaving difficulties at the boundaries both from normal variation, and between different disorders (Kendell & Jablensky, 2003). Mental health research traditionally relies on lengthy structured or semi-structured interviews to provide a "gold standard" highly reproducible syndromic diagnosis (Haro et al., 2006; Rucker et al., 2011), but these are costly to administer, placing a limit on sample sizes.

Common sources of mental disorder status in studies with large sample sizes are symptom scales or check-lists, self-reported clinical diagnoses and medication, and registries. Self-report can be captured in a traditional interview, or using novel forms, such as online questionnaires, which vastly decrease costs of acquiring information (Andersson, Ritterband, & Carlbring, 2008). Registry data no longer comes only from databases set up specifically for research, but can be derived from administrative data. Data-linkage to these sources offers benefits of a wider range of reports without the direct costs of acquiring data, but raises problems of interpreting and validating those reports (Stewart & Davis, 2016).

Clinician diagnoses derived from self-reported or data-linkage, should reflect the outcome of a nuanced clinical assessment, but those people who have received a diagnosis are those who have accessed services, whereas a large proportion of people with a mental disorder are never formally

identified as such (Goldberg & Huxley, 1980). Passage into healthcare will depend upon the type and severity of illness, and patient factors; receiving a diagnosis and treatment depends additionally on clinician and service factors. Such factors are vulnerable to age and cohort effects. For example, antidepressant treatment for those in whom the survey found symptoms of a common mental disorder in the previous year was almost three times more likely in 2009 (15.9%) than it had been in 1993 (5.7%) (Spiers et al., 2016).

A retrospective enquiry adds recall bias for both symptoms and diagnoses. One study estimated that ability to recall a period of sadness likely to represent depression fell from 90% if it occurred in the last year to 41% if it occurred ten years ago (Patten et al., 2012). This problem is not confined to mental health, since self-report of clinician diagnosis of physical disorders including heart failure and previous cancer can be unreliable, leading mostly to under ascertainment (Nord, Mykletun, & Fosså, 2003; Okura, Urban, Mahoney, Jacobsen, & Rodeheffer, 2004). It may be that mental disorders are also under-reported due to perceived stigma of the disorder (Nevin, 2009; Simon & VonKorff, 1995).

Comparison of Approaches in One Resource

UK Biobank (UKB) is a research cohort for which over 503,328 people aged 49-60 enrolled in 2007-2010. This involved questionnaires, biosamples, and consent for linkage of routinely collected healthcare data and to take part in further waves of data collection (Sudlow et al., 2015).

The Mental Health Outcome Consortium was formed to develop mental health phenotyping in UKB. Mental disorder in this context might be both an outcome and a risk factor for other health outcomes. The consortium has focussed on two aspects: validating administrative secondary care diagnostic codes (Davis, Bashford, et al., 2018; Davis, Sudlow, & Hotopf, 2016); and designing an online mental health questionnaire (MHQ) to identify symptom-based outcomes (Davis, Coleman, et al., 2018). Some of the outcomes in the MHQ are based on diagnostic interviews and are analogous to mental disorder diagnoses (e.g. depression and generalised anxiety disorder). Others assess other aspects of mental health such as psychotic experiences (PE) and self-harm. Results of the UKB MHQ are available for 157,366 participants, representing 31% of the original UKB sample (Davis, Coleman, et al., 2018).

UKB now provides multiple indicators that could be used as a means to identify mental disorders, none of which represents a "gold-standard" diagnosis against which the others can be validated. This could lead to confusion and dilemmas as to which measure to use for research. Although there have

been studies that compare individual measures against a conventional gold standard, there are few resources that help guide the choice of imperfect measures in large studies such as UKB. The aim of this study is to compare four indicators of mental health and disorder in UKB for multiple outcomes, in order to guide future research in UKB and the design of similar studies.

Methods

UKB is a major open science resource (Sudlow et al., 2015). Extensive data is already available on the 503,328 volunteers in UKB (UK Biobank, 2018), who responded to invitations sent by mail to people aged 40-69 who lived near to 22 assessment centres in England, Scotland and Wales. The composition has been documented, and it has been noted that the volunteers are not representative of the population as a whole (Davis, Coleman, et al., 2018; Fry et al., 2017), in particularly under-representing people with lower socioeconomic status, people with chronic illness and smokers. This means that the data cannot be used to estimate population prevalence.

The methods used to develop and implement the online MHQ, participation and features of non-participants are described in Davis, Coleman, et al. (2018). All UKB participants with a valid email address were sent a link in 2016-7 (n=339,092), and 46% of those invited submitted valid responses. People who responded had an average age of 65 years and 57% were female. The questionnaire is still open on the website for participants to complete. We report findings based on the dataset released in August 2017 (n=157,363, 31% cohort).

The four main methods of classifying these participants' mental health are: symptom-based outcomes, self-report of diagnosis, hospital data-linkage, and self-report of medication. Brief explanations are provided below, with the full wording, criteria, cut-offs and code lists available in the appendices of supplementary materials. Table 1 shows examples of each method for four outcome groups that will be examined in results. Some of these groups will have more closely aligned concepts that will allow comparison across methods, others will not. For example, psychotic experiences (PE) are not a true ‘symptom’, and most people who have these experiences do not have a psychotic disorder. Therefore self-report diagnosis and hospital data-linkage of psychotic disorder should be viewed as complementary concepts to PE; whereas the depression outcome group are more akin to different methods of ascertaining the same concept.

#Insert table 1 around here

Table 1: Summary of definitions for four measures (columns) that may be used to identify mental health outcomes for four example outcome groups (rows)

Symptom-based outcomes

Lifetime depression, anxiety, bipolar affective disorder (BPAD) and psychotic experiences (PE) make up the lifetime "symptom-based outcomes". Lifetime measures were felt to be important to

generate controls (“never had”) for genetic studies. Depression was assessed using the major depressive disorder section of the Composite International Diagnostic Interview Short Form (CIDI-SF), and anxiety was assessed using the generalized anxiety disorder section of the CIDI-SF, modified to provide lifetime history (Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998; Levinson et al., 2017). There were chosen on the basis of the ability to map on to DSM criteria, results of the validation study carried out by Levinson et al., and to maximise compatibility with studies internationally that were looking at the genetics of depression and anxiety. The CIDI-SF uses DSM-IV criteria, but is also likely to represent a DSM-5 diagnosis as criteria are largely unchanged (American Psychiatric Association, 2013). Further questions assessed probable lifetime history of DSM-defined hypomania/mania; criteria met for at least one week was used as the symptom-based outcome for the BPAD outcome in this study. Lifetime PE, not in themselves a disorder, were assessed using adapted questions from the CIDI (McGrath et al., 2015).

Self-report of diagnosis

Participants were asked about clinician diagnoses of any medical condition at the baseline UKB interview, and were specifically asked about mental disorders in the MHQ. We only use the prompted recall from the MHQ for this analysis. The questionnaire asked participants: "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?" Choices included “depression”, "anxiety, nerves or generalised anxiety disorder", "mania, hypomania, bipolar or manic-depression", "schizophrenia" and "other psychotic illness".

Hospital data linkage

UKB has obtained structured diagnostic information from hospital admissions data to form a virtual hospital registry, combining Hospital Episode Statistics (HES); Scottish Morbidity Record (SMR 1a and 1b); and Patient Episode Database for Wales (PEDW) into a single dataset (UK Biobank, 2014). Dates and completeness of coverage vary: PEDW dates back to 1999, HES to 1997, and SMR to 1981. HES and PEDW have mental health admissions in the same set as general hospital admissions, but Scotland do not. At the time of extraction, the Scottish mental health admissions (SMR-04) were not available in UKB. Therefore participants registered for the UKB in the two Scottish centres were excluded from the comparisons that involve hospital data-linkage, leaving 146,813 participants in England and Wales. HES and PEDW use World Health Organisation (WHO) International Classification of Diseases 10th Revision (ICD-10) to categorise diagnosis (World Health Organization, 1992). Cases were defined as having ever received an ICD-10 diagnosis code relating to depression,

anxiety, BPAD or psychosis (see table 1 or appendix 3 in the supplementary material) in main or any secondary diagnoses. Psychosis codes included depression and BPAD where psychotic symptoms were specified.

Self Report of Medication

At baseline (2007-2010), six to ten years before completion of the MHQ, UKB participants were asked whether they were taking any regular medication, and a nurse interviewer took the names of medication taken. A pre-existing code list of antidepressants, antipsychotics and lithium preparations was used to extract this data (see appendix 4 in supplementary material).

Data and Analysis

The study used the UKB data release application number 16577 (application by GB), including valid MHQ data to June 2017 and hospital admission data 1997-2015, extracted and analysed using R version 3.4.3 (R Core Team, 2017) and code written by JC and KD (Davis, Coleman, et al., 2018). Full data is available from UKB (http://www.ukbiobank.ac.uk/register-apply/).

Confidence intervals are given at 95%, using Wilson's method for proportions. Cohen's kappa was calculated as a measure of agreement between different methods of ascertainment for the same or equivalent outcomes.

Ethical approval

UKB has ethical approval from the North West - Haydock Research Ethics Committee (11/NW/0382) with MHQ approved as an amendment.

Results

Self-reported Diagnosis

Table 2 is a cross-tabulation of overlap between (i) self-reported lifetime diagnosis and (ii) symptom-based outcomes. Percentages refer to the proportion of those with a self-reported diagnosis (row) who met criteria for the specified symptom-based outcomes (column). Of those that reported any mental disorder, 60% also met criteria for any symptom-based outcome, while only 15% of those who reported no mental disorder met any criteria. The self-report status (any vs none) agreed with the symptom-based status in 78%, with a kappa of 0.46. Nearly ninety percent of people reporting BPAD or psychotic disorder met criteria for one or more symptom-based outcome. Regardless which disorder was self-reported, lifetime depression was the most likely symptom-based outcome.

Depression, anxiety and BPAD self-reported diagnoses and symptom-based outcomes are compared in table 3. Depression outcomes had a kappa of 0.46, anxiety outcomes have a kappa of 0.28 and BPAD outcomes have a kappa of 0.24.

Hospital data-linkage

Table 4 shows the partial overlap between the symptom-based outcomes, self-reported diagnosis and hospital data-linkage. The combination of three sources identified depression in 48,794 participants, but the hospital data-linkage only identified 3,034 (6%) of these, most of whom (1,937) were also identified by both of the other two methods. Hospital data-linkage identified 5% of anxiety cases and 9% of BPAD cases identified by any means. Of those with hospital data-linkage diagnosis of psychotic illness (213), the symptom-based outcome of PE was present in 67% (143).

Self-Reported Medication

The snapshot view of selected self-report medication use provided at the baseline assessment is shown in tables 5a-c. Antidepressants were being taken by 8,616 (5.9%) participants, while antipsychotics and lithium were prescribed to less than 500 people each. Eighty-three percent (83%) of all people taking antidepressants were identified as having a lifetime history of depression through one of the three methods. Only half of the participants taking antipsychotics reported PEs or had a diagnosis of psychosis (229/470, 49%), although a further 35% (163/470) had an indicator of affective disorder. Lithium was almost confined to those identified as having an affective disorder – 79% BPAD, 20% depression without evidence of BPAD.

Combinations

Table 4 shows the results of combining symptom-based outcomes, self-reported diagnosis and hospital data-linkage in an additive manner for depression, anxiety and BPAD. In all disorders, symptom-based outcomes, self-report and hospital data-linkage each contribute unique cases – but in different proportions for each disorder.

Combinations of outcomes for the common mental disorders of depression and anxiety are further explored in table SM1 and accompanying text. The symptom-based outcomes were positive for depression or anxiety in 37,629 participants. Self-reported or data-linkage diagnosis of depression or anxiety or self-reported antidepressant medication is positive in 47,321 participants, including 25,920 (55%) who were positive and 21,401 (45%) who were negative on lifetime symptom-based outcomes.

Discussion

In this study we have compared methods of ascertainment for mental health outcomes in UKB from the position that none is equivalent to the outcome of a gold-standard psychiatric interview. This situation is common in large non-specialist research resources, and there is a need for resources to help with decision-making when researchers are faced with a choice of imperfect measures.

We found that the magnitude of the overlap between the measures differed depending on the disorders. Depression outcomes were the most prevalent and had the most overlap between self-report and symptom-based outcomes (kappa=0.46). The proportion of participants with symptom-based outcome who self-reported a diagnosis was 55%, similar to the 61% of people of a similar age in a German study who were positive for lifetime depression on the SCID-I who self-reported a diagnosis (Stuart et al., 2014).

A self-reported diagnosis of "anxiety, nerves or generalised anxiety disorder" had less overlap with the corresponding symptom-based outcome (kappa=0.28), a symptom-based outcome for depression (53%) being more likely than anxiety (26%). Combining depression and generalised anxiety may be an acceptable strategy in population studies, where the concepts are largely overlapping (Gask, Klinkman, Fortes, & Dowrick, 2008), and in our data this led to an improvement in agreement between self-report and symptom-based outcomes over anxiety, but not depression (kappa=0.46).

The conventional models of BPAD, with dramatic and disabling symptoms, would predict a high proportion to have been formally identified, but our symptom-based outcome of BPAD was deliberately fairly wide to facilitate research into the wider spectrum of BPAD (Phillips & Kupfer, 2013), and would include many people who would meet the DSM criteria for BPAD type II as well as BPAD type I. People with BPAD type II will be less likely to be formally diagnosed or require inpatient treatment, and hence will be less commonly identified by a hospital data-linkage. Of those with BPAD symptom-based outcome, 16% self-reported clinician diagnosis and 9% had data-linkage diagnosis. Self-report diagnosis is somewhat higher in this study than in a similar Finnish population study (Perälä et al., 2007) where only 6% of those positive for the CIDI-BPAD outcome self-reported a diagnosis. This may be evidence of a cohort effect of different diagnostic behaviour or patient awareness between countries or over time.

PE and psychotic disorder are not equivalent, but complementary categories. We found that PE was almost ten times more common than psychotic disorder reported by the participant and/or hospital data-linkage (prevalence of PE 4.7% vs psychotic disorder diagnosis 0.5%). The Finnish study (Perälä et al., 2007) found the rates of PE and psychosis diagnosis to be 3.0% and 3.3% respectively. The lower prevalence of PE may be partly due to the mode of administration being interview, as PE are more likely to be endorsed in self-completed measures (Linscott & Van Os, 2013). The higher levels of diagnosis of a psychosis diagnosis may be partly because the registry used in the Finnish study goes back further in time, but may also be related to participation bias. The Finnish study was a modest size study aiming at representativeness, with a participation rate of 93% of those selected, whereas UKB followed a different model, requesting volunteers from the community (Davis, Coleman, et al., 2018; Fry et al., 2017): people with an enduring psychotic disorder may have been less willing and/or able to volunteer.

Of the three self-reported medication classes investigated, antidepressants were the most commonly reported. Even so, antidepressant prescription could only identify 15-17% of people with those symptom-based outcomes of depression and anxiety. This is inevitable given the snapshot nature of the ascertainment of medication, the "treatment gap" (Kohn, Saxena, Levav, & Saraceno, 2004), and appropriate management of lifetime mental disorder without medication. Surprisingly, only 49% of those taking antipsychotics were positive on a measure of PE or psychosis, 35% had an affective disorder and 13% neither. This fits with literature on the extended and off-label prescribing of antipsychotics (Carton et al., 2015; Pringsheim, Gardner, & Patten, 2015).

Method of ascertainment

Symptom-based outcomes do not require participants to have accessed care to detect a disorder, making them potentially the most sensitive out of the measures we compared, although the retrospective nature is likely to reduce sensitivity for distant episodes. By analysing participant responses to particular questions, it may also be possible to also look at subtypes or specific phenotypes or manipulate thresholds. Symptoms were collected using CIDI-SF modules. The CIDI was created for the World Health Organisation (WHO) programme, and supported by them, although the short form is not currently supported by the WHO. Such measures are popular in surveys (McDowell, 2006; van Ballegooijen, Riper, Cuijpers, van Oppen, & Smit, 2016), although they can be over-inclusive as they lack the ability to rule out other causes of the same symptoms (e.g. thyroid disturbance mimicking anxiety). Alternatives to the CIDI-SF may have different, possibly better, performance – but this has not been tested.

Administration of self-report diagnostic scales online is now an established practice (Andersson et al., 2008; Nguyen, Klein, Meyer, Austin, & Abbott, 2015), but there is generally less validation data available for measures administered electronically or via the internet (Buchanan, 2003; van Ballegooijen et al., 2016). The performance of the CIDI-SF modules that were administered in the online MHQ have been positively validated in at least two independent studies (Carlbring et al., 2002; Levinson et al., 2017).

Self-reported clinician diagnosis is an easily obtainable measure, which allowed the MHQ to ask about a wide range of outcomes. As predicted, the diagnosis prevalence was lower than the symptom-based outcome prevalence in the MHQ in most categories. The exception was generalised anxiety – which may be related to the wording of the question regarding anxiety diagnosis being vague. The presence of self-reported diagnosis was associated with a greater risk of all symptom-based outcomes, not just for equivalent outcomes, which reflects the comorbidities between disorders often unrecognised (Oiesvold et al., 2013; Whiteford et al., 2015). Another source of self-reported diagnosis in UKB are those reported during the baseline assessment. On that occasion, participants were not prompted to recall specific diagnoses, and had to disclose them face-to-face. The prevalence of self-reported mental prevalence was lower on that occasion, with depression reported by only 6.5%, as opposed to 21% at the MHQ. This is likely to do with the prompted recall, but may also be due to stigma during a face to face interview and new diagnoses since baseline.

The hospital data-linkage provided by UKB leverages national statistics to identify outcomes that are commonly documented in hospital admissions. The nature and patient pathway of mental disorders mean only the most severe cases are likely to be the cause of an admission (Goldberg & Huxley, 1980). Moreover, these episodes may have happened many decades ago, before 1997 when the data for England starts. Most mentions of mental disorder will therefore be secondary diagnoses in participants admitted to hospital with other problems, which have not been specifically validated (Davis, Bashford, et al., 2018). In this study, the low numbers identified in hospital data-linkage, with high levels of lifetime symptom-based outcomes in those individuals, suggests a specific but insensitive measure. Registries based on data-linkage to outpatient attendance or primary care consultations may give a more sensitive measure, although it is likely to be more complex to define cases given the myriad of coding types in these records (John et al., 2016; Spiranovic, Matthews, Scanlan, & Kirkby, 2016).

The use of self-reported medication data is potentially problematic. Bias in recall of medication is very common, perhaps more so in psychotropics (Gnjidic, Du, Pearson, Hilmer, & Banks, 2017). Objective ascertainment of prescribed medication is likely to be provided in the future by linkage to primary care data, and in some studies, pharmacy claims data has been successfully used to supplement self-reported medication (Drieling et al., 2016; Gnjidic et al., 2017). However, there will remain the likelihood that medication will have poor sensitivity for case finding in mental health, as psychotropics will never be prescribed to all of those with a lifetime history, and poor specificity as they are prescribed for many things outside of mental health. In the case of using medication in the UKB to supplement MHQ findings, there is the added problem of the snapshot of medication taken being ascertained around seven years prior to the MHQ administration, and therefore being unable to reflect new-onset disorders and prescriptions.

Algorithmic approaches can be taken that exploit the strengths of each measure to produce a compound measure. Algorithms will include combining cases from two or more outcome types as done for this genomic study of depression in UKB using baseline self-report and hospital diagnosis (Howard et al., 2018). Items can also be grouped into new criteria as was done to define mood disorders at baseline (Smith et al., 2013). Another approach, previously suggested in the case-control definitions defined by the UKB mental health outcomes group, uses symptom-based outcomes for cases, but exclude from controls those who self-reported diagnosis or had data-linkage diagnosis or suggestive medication. Taking the BPAD row from table 4 as a simplified example: 2,247 people were positive for the symptom-based outcome and 155,119 were negative; out of those who did not meet criteria, 177 had a hospital diagnosis of BPAD, 326 more reported a diagnosis of BPAD; and 35 more reported taking lithium (table 5c)– all of these are suggestive of BPAD. To minimise false-positives and false-negatives in the BPAD item, these 538 suggestive participants can be excluded from cases and controls, leaving 2,247 cases and 154,581 controls. Further algorithms incorporating hospital and primary care data for severe mental illness and common mental disorder in the full cohort are due to be published by UKB in 2019-20 – as has already been done for stroke and myocardial infarction.

Does it matter?

We have shown that different methods of ascertainment of mental disorder can result in different groups of participants being identified as cases. This poor agreement between methods of ascertainment could be problematic for research consistency and reproducibility. However, there is evidence that even with poor agreement at the level of disorder diagnosis, there can be similarity at

the biological level. For example, a twin study (Torvik et al., 2018) reported that cases derived from interview diagnoses had limited overlap with those selected by data-linkage (primary and secondary) – for depression 36% interview positive were also on primary care registry, while 48% of those in registry were interview positive, with less overlap for anxiety (21%/46%) and alcohol use disorder (3%/33%). Despite this, the genetic features identified in the interview and registry groups were highly correlated within each diagnosis, approaching unity for depression and anxiety disorders. It remains to be seen whether the same will be true for the different cohorts selected in UKB – certainly focussing exclusively on very highly selected outcomes such as hospital data-linkage means including biases to do with health service utilisation that may not relate to underlying mental health need (Roberts et al., 2018).

Genome-wide association studies (GWAS) often pool cases and controls from different cohorts. Studies that define DSM disorders using clinical interview, self-report diagnosis, symptom-based outcomes, or combinations thereof might be combined in order to achieve the necessary size of sample. The results will then depend heavily on whether the biology converges on a single disorder or converges on the different definitions (Vrieze, Iacono, & McGue, 2012). A massed GWAS of depression (Wray et al., 2018) included cases that were defined at interview (PGC29, GenScot), treatment registers (iPSYCHE, GERA), self-report diagnosis (23andMe) and a combination (DeCODE, UKB [prior to MHQ results]) showed strong genetic correlation between the studies. The combined GWAS also showed enrichment of the targets of antidepressant treatment. These results suggest that weakening the phenotype can reveal interesting and relevant biology.

On some occasions, we have found that different measures have indicated different disorders for the same individuals, which could lead to confusion in research concentrated on a narrowly defined diagnosis. However, this reflects established findings of a high degree of comorbidity and cross-over in mental disorders (Davis, Bashford, et al., 2018; First, 2005; Gask et al., 2008), probably due to shared etiology and pathology (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Elliott, Romer, Knodt, & Hariri, 2018) that is poorly translated into categorical diagnostic classifications. Other models for understanding mental disorder have been suggested, and some of these ideas could be translated to measures for research in large cohorts (Carcone & Ruocco, 2017; Clark et al., 2017; Vrieze et al., 2012), but diagnostic categories continue to be utilised widely.

Implications

For users of UKB, the symptom-based outcomes defined in the MHQ offers advantages: they will select a large proportion of the participants with a likely disorder; many have been validated externally; and there is scope to customise, such as for different thresholds. However, self-report, hospital data-linkage and medication may also be able to identify unique cases, and may have high predictive validity. In some cases, it would seem sensible to add cases together. Another approach is to use the symptom-based outcome to define the cases, and define the controls to exclude positives on the other measures. For some questions, the sample and measures in the MHQ may be too limiting, and unprompted baseline self-report supplemented by hospital data-linkage will have to be used (Howard et al., 2018), which are highly selected, until primary care data and algorithms are released. Comorbidity between mental disorders is high, and interpretation of this may need consideration. Given the high degree of flexibility that UKB affords, researchers should consider the breadth and granularity of the mental health diagnosis needed alongside the consideration of the variables used to define them, so that the most appropriate combination of measure and outcome can be chosen to best address the research question.

Other studies could learn from the experience in UKB in three main ways. Firstly, under-recognition, fluctuating course and self-management of most mental disorder means questions about lifetime symptoms are needed to identify those who have never had a disorder. Second, comorbidity between the mental disorders is high, and this needs to be acknowledged in the design and interpretation of mental health questionnaires. Thirdly, registries, data-linkage and measures of treatment will underestimate numbers of cases of mental disorder, but do provide further information.

Strengths and weaknesses

UKB aims to produce and adjudicate outcomes in a clear, expert-led manner. The Mental Health Outcomes Consortium has worked with UKB to implement the MHQ, and the present analysis was planned to clarify the different mental health definitions now present in UKB.

The MHQ had a very good response rate compared with previous UKB online questionnaires, and it gives an unparalleled sample size for a mental health survey. However, like much observational research, it is subject to participation bias in its volunteers (Davis, Coleman, et al., 2018; Fry et al., 2017). Given that participation in research can be patterned by mental health (Atherton, Fuller, Shepherd, Strachan, & Power, 2008; Knudsen, Hotopf, Skogen, Øverland, & Mykletun, 2010), it may be that people with severe symptoms of mental disorder were less likely to volunteer or complete

the MHQ, as might be suggested by the small number of people with a hospital data-linkage diagnosis of a psychotic disorder, which may limit generalisability of our findings to other settings.

The measures in the MHQ were felt to be the most suitable for defining lifetime mental disorders within the constraints of a short survey and maintaining compatibility with existing genetic studies. The online CIDI-SF has been validated, but only for depression in the lifetime version. The questions used to assess for symptoms of mania / hypomania have not been externally validated. For both instruments, it is likely that the lifetime version is affected by recall bias. Further, the UKB data-linkage and medication aspects are currently limited. Hospital admission data will capture few with mental disorders, so we will welcome the forthcoming linkages to primary care data. Medication was self-reported and on a single occasion that was seven to ten years prior to the symptom-based outcome: again it may be better after linkage to primary care data.

Conclusions.

Large cohort studies provide great potential for interesting discovery, but using these datasets involves confronting problems with definitions of disorders, data quality and incomplete coverage. Mental health research is further hampered the challenge that many mental disorders are under-recognised and under-represented in healthcare data. UKB is a rich observational resource due to its size, extensive baseline measures and linkages to national administrative records. The utility of UKB for mental health research has been improved by the UKB MHQ. We have shown that, in general, the numbers of cases identified by lifetime symptom-based diagnosis exceeds those identified with self-report diagnosis, hospital data-linkage and psychotropic medication, with an overlap between measures that differs between the disorders under study. The advantage of symptom-based lifetime classification of mental disorder is sensitivity across the severity spectrum, and many of the symptom-based outcomes have been validated against psychiatric interview elsewhere. However, other mental health ascertainment methods could complement symptom-based outcome measures in research. UKB and other open science projects lend themselves to innovative, well-described and reported approaches that can be scrutinised by the community. The ideas and results of this exploratory analysis highlight the strengths and limitations of both the indicators in large cohort studies, and the mental disorder diagnosis itself, which we hope will assist those planning to address the important questions in mental health and wider research.

ix)

American Psychiatric Association. (2013). Highlights of Changes from DSM-IV-TR to DSM-5 (https://dsm.psychiatryonline.org/doi/full/10.1176/appi.books.9780890425596.changes). In APA (Ed.), *Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition*: American Psychiatric Association Publishing.

Andersson, G., Ritterband, L. M., & Carlbring, P. (2008). Primer for the assessment, diagnosis and delivery of Internet interventions for (mainly) panic disorder. Lessons learned from our research groups. *Clinical Psychologist, 12*(1), 1-8. doi:10.1080/13284200802069027

Atherton, K., Fuller, E., Shepherd, P., Strachan, D. P., & Power, C. (2008). Loss and representativeness in a biomedical survey at age 45 years: 1958 British birth cohort. *Journal of epidemiology and community health, 62*(3), 216-223. doi:10.1136/jech.2006.058966

Buchanan, T. (2003). Internet-based Questionnaire Assessment: Appropriate Use in Clinical Contexts. *Cognitive Behaviour Therapy, 32*(3), 100-109. doi:10.1080/16506070310000957

Carcone, D., & Ruocco, A. C. (2017). Six years of research on the National Institute of Mental Health's research domain criteria (RDoC) initiative: a systematic review. *Frontiers in cellular neuroscience, 11*, 46.

Carlbring, P., Forslin, P., Ljungstrand, P., Willebrand, M., Strandlund, C., Ekselius, L., & Andersson, G. (2002). Is the Internet-administered CIDI-SF Equivalent to a Clinician-administered SCID Interview? *Cognitive Behaviour Therapy, 31*(4), 183-189. doi:10.1080/165060702321138573

Carton, L., Cottencin, O., Lapeyre-Mestre, M., A. Geoffroy, P., Favre, J., Simon, N., . . . Rolland, B. (2015). Off-Label Prescribing of Antipsychotics in Adults, Children and Elderly Individuals: A Systematic Review of Recent Prescription Trends. *Current Pharmaceutical Design, 21*(23), 3280-3297.

Casey, & Kelly, B. (2007). Classification of Psychiatric Disorders. In F. J. Fish, P. Casey, & B. Kelly (Eds.), *Fish's clinical psychopathology: signs and symptoms in psychiatry* (pp. 1-13). London: RCPsych Publications.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest, 18*(2), 72-145. doi:10.1177/1529100617727266

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet, 381*(9875), 1371-1379.

Davis, K. A. S., Bashford, O., Jewell, A., Shetty, H., Stewart, R. J., Sudlow, C. L. M., & Hotopf, M. (2018). Using data linkage to electronic patient records to assess the validity of selected mental health diagnoses in English Hospital Episode Statistics (HES). *PLoS One, doi:10.1371/journal.pone.0195002*.

Davis, K. A. S., Coleman, J., Adams, M., Allen, N., Breen, G., Cullen, B., . . . Hotopf, M. (2018). Mental Health in UK Biobank – development, implementation and results from an online questionnaire completed by 157,366 participants. *BJPsych Open, 4*, 83-90. doi:10.1192/bjo.2018.12

Davis, K. A. S., Sudlow, C. L., & Hotopf, M. (2016). Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. *BMC psychiatry, 16*(1), 263.

Drieling, R. L., LaCroix, A. Z., Beresford, S. A., Boudreau, D. M., Kooperberg, C., & Heckbert, S. R. (2016). Validity of Self-Reported Medication Use Compared With Pharmacy Records in a Cohort of Older Women: Findings From the Women's Health Initiative. *Am J Epidemiol, 184*(3), 233-238. doi:10.1093/aje/kwv446

Elliott, M. L., Romer, A., Knodt, A. R., & Hariri, A. R. (2018). A Connectome-wide Functional Signature of Transdiagnostic Risk for Mental Illness. *Biological Psychiatry*. doi:10.1016/j.biopsych.2018.03.012

First, M. B. (2005). Mutually exclusive versus co-occurring diagnostic categories: the challenge of diagnostic comorbidity. *Psychopathology, 38*(4), 206-210.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., . . . Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with the general population. *Am J Epidemiol*. doi:10.1093/aje/kwx246

Gask, L., Klinkman, M., Fortes, S., & Dowrick, C. (2008). Capturing complexity: the case for a new classification system for mental disorders in primary care. *European Psychiatry, 23*(7), 469-476.

Gnjidic, D., Du, W., Pearson, S. A., Hilmer, S. N., & Banks, E. (2017). Ascertainment of self-reported prescription medication use compared with pharmaceutical claims data. *Public Health Res Pract, 27*(4). doi:10.17061/phrp27341702

Goldberg, D. P., & Huxley, P. (1980). *Mental illness in the community: the pathway to psychiatric care*: Tavistock Publications Limited (Republished by Routledge 2011).

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R., . . . Kessler, R. C. (2006). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *Int J Methods Psychiatr Res, 15*(4), 167-180.

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., . . . McIntosh, A. M. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature communications, 9*(1), 1470. doi:10.1038/s41467-018-03819-3

John, A., McGregor, J., Fone, D., Dunstan, F., Cornish, R., Lyons, R. A., & Lloyd, K. R. (2016). Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Medical Informatics and Decision Making, 16*(1), 35. doi:10.1186/s12911-016-0274-7

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry, 160*(1), 4-12.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H. U. (1998). The World Health Organization composite international diagnostic interview short-form (CIDI-SF). *Int J Methods Psychiatr Res, 7*(4), 171-185.

Knudsen, A. K., Hotopf, M., Skogen, J. C., Øverland, S., & Mykletun, A. (2010). The health status of nonparticipants in a population-based health study: The Hordaland Health Study. *Am J Epidemiol, 172*(11), 1306-1314. doi:10.1093/aje/kwq257

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World health Organization, 82*(11), 858-866.

Levinson, D., Potash, J., Mostafavi, S., Battle, A., Zhu, X., & Weissman, M. (2017). Brief Assessment Of Major Depression For Genetic Studies: Validation Of CIDI-SF Screening With SCID Interviews. *European Neuropsychopharmacology, 27*, S448. doi:10.1016/j.euroneuro.2016.09.514

Linscott, R., & Van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol Med, 43*(6), 1133-1149.

McDowell, I. (2006). Anxiety & Depression. In *Measuring health: a guide to rating scales and questionnaires* (pp. 273-393): Oxford University Press.

McGrath, J. J., Saha, S., Al-Hamzawi, A., Alonso, J., Bromet, E. J., Bruffaerts, R. … Fayyad, J. (2015). Psychotic experiences in the general population: a cross-national analysis based on 31 261 respondents from 18 countries. *JAMA Psychiatry, 72*(7), 697-705.

Nevin, R. L. (2009). Low validity of self-report in identifying recent mental health diagnosis among US service members completing Pre-Deployment Health Assessment (PreDHA) and deployed to Afghanistan, 2007: a retrospective cohort study. *BMC Public Health, 9*(1), 376.

Nguyen, D. P., Klein, B., Meyer, D., Austin, D. W., & Abbott, J.-A. M. (2015). The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *Journal of medical Internet research, 17*(9), e218-e218. doi:10.2196/jmir.4195

Nord, C., Mykletun, A., & Fosså, S. D. (2003). Cancer patients' awareness about their diagnosis: a population-based study. *Journal of Public Health, 25*(4), 313-317. doi:10.1093/pubmed/fdg076

Oiesvold, T., Nivison, M., Hansen, V., Skre, I., Ostensen, L., & Sorgaard, K. (2013). Diagnosing comorbidity in psychiatric hospital: challenging the validity of administrative registers. *BMC psychiatry, 13*(1), 13.

Okura, Y., Urban, L. H., Mahoney, D. W., Jacobsen, S. J., & Rodeheffer, R. J. (2004). Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *Journal of clinical epidemiology, 57*(10), 1096-1103. doi:10.1016/j.jclinepi.2004.04.005

Patten, S. B., Williams, J. V. A., Lavorato, D. H., Bulloch, A. G. M., D'Arcy, C., & Streiner, D. L. (2012). Recall of recent and more remote depressive episodes in a prospective cohort study. *Social Psychiatry and Psychiatric Epidemiology, 47*(5), 691-696. doi:10.1007/s00127-011-0385-5

Perälä, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsä, E., Pirkola, S., . . . Kieseppä, T. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry, 64*(1), 19-28.

Phillips, M. L., & Kupfer, D. J. (2013). Bipolar disorder diagnosis: challenges and future directions. *The Lancet, 381*(9878), 1663-1671. doi:10.1016/S0140-6736(13)60989-7

Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *Lancet, 370*(9590), 859-877.

Pringsheim, T., Gardner, D., & Patten, S. B. (2015). Adjunctive treatment with quetiapine for major depressive disorder: are the benefits of treatment worth the risks? *BMJ : British Medical Journal, 350*. doi:10.1136/bmj.h569

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.3). https://www.R-project.org/: R Foundation for Statistical Computing, Vienna, Austria.

Roberts, T., Miguel Esponda, G., Krupchanka, D., Shidhaye, R., Patel, V., & Rathod, S. (2018). Factors associated with health service utilisation for common mental disorders: a systematic review. *BMC psychiatry, 18*(1), 262. doi:10.1186/s12888-018-1837-1

Rucker, J., Newman, S., Gray, J., Gunasinghe, C., Broadbent, M., Brittain, P., . . . Stewart, R. (2011). OPCRIT+: an electronic system for psychiatric diagnosis and data collection in clinical and research settings. *The British Journal of Psychiatry, 199*(2), 151-155.

Simon, G. E., & VonKorff, M. (1995). Recall of psychiatric history in cross-sectional surveys: implications for epidemiologic research. *Epidemiol Rev, 17*(1), 221-227.

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., . . . Mackay, D. (2013). Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One, 8*(11), e75362.

Spiers, N., Qassem, T., Bebbington, P., McManus, S., King, M., Jenkins, R., . . . Brugha, T. S. (2016). Prevalence and treatment of common mental disorders in the English national population, 1993–2007. *The British Journal of Psychiatry, 209*(2), 150-156.

Spiranovic, C., Matthews, A., Scanlan, J., & Kirkby, K. C. (2016). Increasing knowledge of mental illness through secondary research of electronic health records: opportunities and challenges. *Advances in Mental Health, 14*(1), 14-25. doi:10.1080/18387357.2015.1063635

Stewart, R., & Davis, K. (2016). 'Big data' in mental health research: current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology, 51*(8), 1055-1072. doi:10.1007/s00127-016-1266-8

Stuart, A. L., Pasco, J. A., Jacka, F. N., Brennan, S. L., Berk, M., & Williams, L. J. (2014). Comparison of self-report and structured clinical interview in the identification of depression. *Comprehensive Psychiatry, 55*(4), 866-869. doi:10.1016/j.comppsych.2013.12.019

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine, 12*(3), e1001779.

Torvik, F. A., Ystrom, E., Gustavson, K., Rosenström, T. H., Bramness, J. G., Gillespie, N., . . . Reichborn-Kjennerud, T. (2018). Diagnostic and genetic overlap of three common mental disorders in structured interviews and health registries. *Acta Psychiatrica Scandinavica, 137*(1), 54-64. doi:10.1111/acps.12829

UK Biobank. (2014). *Mapping inpatient hospital data across England, Scotland and Wales*. Retrieved from http://biobank.ctsu.ox.ac.uk/crystal/docs/inpatient_mapping.pdf (accessed 04 April 2019)

UK Biobank. (2018). UK Biobank Data Showcase. Retrieved from http://biobank.ctsu.ox.ac.uk/crystal/ (accessed 04 April 2019)

van Ballegooijen, W., Riper, H., Cuijpers, P., van Oppen, P., & Smit, J. H. (2016). Validation of online psychometric instruments for common mental health disorders: a systematic review. *BMC psychiatry, 16*(1), 45. doi:10.1186/s12888-016-0735-7

Vrieze, S. I., Iacono, W. G., & McGue, M. (2012). Confluence of genes, environment, development, and behavior in a post Genome-Wide Association Study world. *Dev Psychopathol, 24*(4), 1195-1214. doi:10.1017/s0954579412000648

Whiteford, H. A., Ferrari, A. J., Degenhardt, L., Feigin, V., & Vos, T. (2015). The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One, 10*(2).

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. http://www.who.int/classifications/icd/icdonlineversions/en/ (accessed 04 April 2019)

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . The Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*(5), 668-681. doi:10.1038/s41588-018-0090-3

x)

Table1: Summary of definitions for four measures (columns) that may be used to identify mental health outcomes for five example outcome groups (rows)

	Symptom-based outcome (see also appendix 2)	**Self-report diagnosis**	**Hospital data-linkage 1997-2015 (see also appendix 3)**	**Self-report medication 2007-10 (see also appendix 4)**
Depression outcomes	Positive for major depressive disorder ever in MHQ (CIDI-SF lifetime). Prevalence 24%	Endorsed clinician diagnosis of "depression" in MHQ. Prevalence 21%	Diagnosis of ICD-10 depressive disorder (F32-33) on inpatient record. Prevalence 2%	Reported use of an antidepressant (prevalence 5%), antipsychotic (prevalence 0.3%) or lithium (prevalence 0.1%) at baseline.
Anxiety outcomes	Positive for generalised anxiety disorder ever in MHQ (CIDI-SF lifetime). Prevalence 7%	Endorsed clinician diagnosis of "anxiety, nerves or generalised anxiety disorder" in MHQ. Prevalence 14%	Diagnosis of ICD-10 neurotic disorders (F4x) on inpatient record. Prevalence 1%	Reported use of an antidepressant at baseline. Prevalence 5%
Bipolar affective disorder (BPAD) outcomes	Positive for wider bipolar criteria ever in MHQ (reflecting DSM IV hypomania/mania criteria). Prevalence 2%	Endorsed clinician diagnosis of "mania, hypomania, bipolar or manic-depression" in MHQ. Prevalence 1%	Diagnosis of ICD-10 mania or BPAD (F30-31) on inpatient record. Prevalence 0.2%	Reported use of lithium (prevalence 0.1%) or an antipsychotic (prevalence 0.3%) at baseline.
Psychotic experience (PE) outcomes	Endorsed one or more of four PEs ever (adapted CIDI PE lifetime)*. Prevalence 5%	Endorsed clinician diagnosis of "schizophrenia" or "other psychotic illness" in MHQ. Prevalence 1%	Diagnosis of ICD-10 schizophrenia spectrum (F2x) or affective psychosis (F30.2, F31.2, F31.5, F32.3, F33.3) on inpatient record. Prevalence 0.1%	Reported use of antipsychotic at baseline. Prevalence 0.3%

Footnotes

*PE are not true 'symptoms' but outcome that can be related to psychotic disorder

BPAD: Bipolar affective disorder; CIDI-SF: Composite International Diagnostic Interview Short Form; ICD-10: International classification of diseases; PE: Psychotic experience

Prevalence refers to criteria positive in this sample of 157,363 UKB volunteers who completed the MHQ.

Tables 2: Symptom-based outcomes (SBO, columns) and self-reported diagnoses (SR, rows). Numbers define participants with both stated symptom-based outcome and self-report (SBO ∩ SR) and % is proportion of participants with given self-report also having given symptom-based outcome (SBO|SR).

		Overall		**Symptom-based outcome (SBO)** n. SBO∩SR (SBO|SR %)					
		n.	prev. in sample	Depression	Anxiety	Wide bipolar definition	PE*	Any SBO	No SBO
Overall	n.	157363	na	37434	11111	2396	7803	44598	112765
	prev. in sample	na	na	24%	7%	2%	5%	28%	72%
Self-report diagnosis (SR)	Depression	33424	21%	20714 (62%)	7173 (21%)	1314 (4%)	3239 (10%)	22651 (68%)	10773 (32%)
	Anxiety	22036	14%	11632 (53%)	5711 (26%)	813 (4%)	2051 (9%)	13365 (61%)	8670 (39%)
	BPAD	837	1%	599 (72%)	248 (30%)	391 (47%)	358 (43%)	737 (88%)	100 (12%)
	Psychosis	723	1%	491 (68%)	247 (34%)	187 (26%)	458 (63%)	635 (88%)	88 (12%)
	Panic disorder	8704	6%	4555 (52%)	2424 (28%)	399 (5%)	1024 (12%)	5273 (61%)	3431 (39%)
	Eating disorder	1851	1%	1048 (57%)	495 (27%)	101 (5%)	279 (15%)	1201 (65%)	650 (35%)
	Personality disorder	385	<1%	270 (70%)	171 (44%)	63 (16%)	141 (37%)	324 (84%)	61 (16%)
	Any self-report	48230	31%	25495 (53%)	9081 (19%)	1721 (4%)	4255 (9%)	28739 (60%)	19491 (40%)
	No self-report	109133	69%	11938 (11%)	2030 (2%)	675 (1%)	3548 (3%)	15859 (15%)	93274 (85%)

Footnotes: BPAD = bipolar affective disorder, PE = psychotic experience

*PE are not true 'symptoms' but outcome that can be related to psychotic disorder

For definitions of symptom-based-outcomes, please see appendix 2 in supplementary material.

Table 3: The overlap of self-report (A) and symptom-based outcome (B) for selected diagnoses, showing the intersect (A ∩ B), proportion overlap (B|A & A|B) and agreement (kappa).

	n. Self-report (A)	**n. Symptom-based outcome (B)**	**n. Self-report AND Symptom-based outcome (A∩B)**	**% Symptom-based outcome given Self-report (B | A)**	**% Self-report given Symptom-based outcome (A | B)**	**kappa**
Depression	33424	37434	20714	62%	55%	0.46
Anxiety	22036	11111	5711	26%	51%	0.28
BPAD	837	2396	391	47%	16%	0.24

Footnotes: BPAD = bipolar affective disorder

Table 4: Identification of five mental health outcomes using symptom-based outcomes, self-report diagnosis and hospital data-linkage, for participants from England and Wales (n=146,813).

	Any	**Symptom criteria (a)**		**Self-report (b)**		**Hospital data-linkage (c)**		**Combinations**			
		Total	Alone	Total	Alone	Total	Alone	a∩b	a∩c	b∩c	all three
Depression	48794	35140 (72%)	15472 (32%)	31381 (64%)	11919 (24%)	3034 (6%)	257 (1%)	19462 (40%)	2143 (4%)	2571 (5%)	1937 (4%)
Anxiety	35136	16806 (48%)	8324 (24%)	26124 (74%)	17264 (49%)	1770 (5%)	555 (2%)	8349 (24%)	704 (2%)	571 (2%)	571 (2%)
BPAD	2709	2247 (83%)	1875 (69%)	783 (29%)	337 (12%)	245 (9%)	37 (1%)	364 (13%)	194 (7%)	120 (4%)	112 (4%)
PE*	7686	7390 (96%)	6920 (90%)	684 (9%)	226 (3%)	213 (3%)	46 (1%)	434 (6%)	143 (2%)	131 (2%)	107 (1%)

Footnotes: See table 1 and appendices for definitions. BPAD = bipolar affective disorder, PE = psychotic experience

*PE are not true 'symptoms' but outcome that can be related to psychotic disorder. Self-report and hospital data-linkage, in contrast, represent psychotic disorders.

Total = n. participants positive on given measure for given outcome (% positive for measure / positive for outcome).

Alone = n. participants that were positive for given measure and not for other measures in given outcome (% positive for this measure alone / positive for outcome)

Combinations: x∩y = participants positive for both given criteria, irrespective of whether positive for third

Table 5a-c: Self-reported psychotropic use at baseline against psychiatric indication by three criteria: symptom-based outcome, self-report diagnosis and hospital data-linkage. % = proportion of cases screening positive for each criteria who reported medication use, except bottom row. Bottom row shows proportion of all participants reporting medication use who screened positive for each disorder.

(a) Self-report of any antidepressant for participants with depression and anxiety outcomes.

	Depression	**Anxiety**	**Nil**
Symptom-based outcome	5352/35140 (15.2%)	2355/10415 (22.6%)	
Self report diagnosis	6378/31381 (20.3%)	4427/26124 (16.9%)	
Hospital data-linkage	1492/2858 (52.2%)	533/1770 (30.1%)	
Self-report antidepressant given above criteria	7137/47278 (15.1%)	5123/31071 (16.5%) *excluding depression* 556/10829 (5.1%)	923/88706 (1.0%)
Any criteria given self-report antidepressant	7137/8616 (82.8%)	*excluding depression* 556/8616 (6.5%)	923/8616 (10.7%)

Footnotes: See table 1 and appendices for definitions.

(b) Self-report of any antipsychotic for participants with psychotic experiences or psychotic disorder (PE), BPAD and depression outcomes.

	PE*	**BPAD**	**Depression**	**Nil**
Symptom-based outcome	203/7390 (2.7%)	103/2247 (4.6%)	300/35140 (0.9%)	
Self report	163/684 (23.8%)	135/783 (17.2%)	277/31381 (0.9%)	
Hospital data-linkage	84/213 (39.4%)	68/245 (27.8%)	105/2858 (3.7%)	
Self-report antipsychotic given above criteria	229/7686 (3.1%)	161/2709 (5.9%) *excluding PE* 42/1890 (2.2%)	354/47278 (0.7%) *excluding PE and BPAD* 121/41359 (0.3%)	78/95879 (0.1%)
Any criteria given self-report antipsychotic	229/470 (48.7%)	*excluding PE* 42/470 (8.9%)	*excluding PE and BPAD* 121/470 (25.7%)	78/470 (16.6%)

Footnotes: See table 1 and appendices for definitions. BPAD = bipolar affective disorder, PE = psychotic experience

*PE are not true 'symptoms' but outcome that can be related to psychotic disorder. Self-report and hospital data-linkage, in contrast, represent psychotic disorders.

(c) Self-report of lithium prescription for participants with BPAD and depression outcomes.

	BPAD	Depression	Nil
Symptom-based outcome	73/2247 (3.2%)	127/35140 (0.4%)	na
Self report	119/783 (15.2%)	111/31381 (0.4%)	na
Hospital data-linkage	67/245 (27.3%)	50/2858 (1.7%)	na
Self-report lithium given above criteria	131/2709 (4.8%)	146/47278 (0.3%) *excluding BPAD* 34/45195 (0.1%)	1/98909 (0.0%)
Any criteria given self-report lithium	131/166 (78.9%)	*excluding BPAD* 34/166 (20.5%)	1/166 (0.6%)

Footnotes: See table 1 and appendices for definitions. BPAD = bipolar affective disorder

xi) Figure headers

Nil

xii) Appendices

Table SM1: Overlap of routine items for common mental disorder and symptom-based outcome for common mental disorder

Appendix 1: Questionnaire wording and format

Appendix 2: Case Criteria Derived from the UK Biobank Mental Health Questionnaire

Appendix 3: ICD-10 codes used for hospital data-linkage

Appendix 4: UKB medication codes used