CrossMark

Latent subtypes of manic and/or irritable episode symptoms in two population-based cohorts

Ryan Arathimos, Chiara Fabbri, Evangelos Vassos, Katrina A. S. Davis, Oliver Pain, Alexandra Gillett, Jonathan R. I. Coleman, Ken Hanscombe, Saskia Hagenaars, Bradley Jermy, Anne Corbett, Clive Ballard, Dag Aarsland, Byron Creese and Cathryn M. Lewis

Background

Mood disorders are characterised by pronounced symptom heterogeneity, which presents a substantial challenge both to clinical practice and research. Identification of subgroups of individuals with homogeneous symptom profiles that cut across current diagnostic categories could provide insights in to the transdiagnostic relevance of individual symptoms, which current categorical diagnostic systems cannot impart.

Aims

To identify groups of people with homogeneous clinical characteristics, using symptoms of manic and/or irritable mood, and explore differences between groups in diagnoses, functional outcomes and genetic liability.

Method

We used latent class analysis on eight binary self-reported symptoms of manic and irritable mood in the UK Biobank and PROTECT studies, to investigate how individuals formed latent subgroups. We tested associations between the latent classes and diagnoses of psychiatric disorders, sociodemographic characteristics and polygenic risk scores.

Results

Five latent classes were derived in UK Biobank (N = 42 183) and were replicated in the independent PROTECT cohort (N = 4445), including 'minimally affected', 'inactive restless', active restless',

'focused creative' and 'extensively affected' individuals. These classes differed in disorder risk, polygenic risk score and functional outcomes. One class that experienced disruptive episodes of mostly irritable mood largely comprised cases of depression/anxiety, and a class of individuals with increased confidence/creativity reported comparatively lower disruptiveness and functional impairment.

Conclusions

Findings suggest that data-driven investigations of psychopathological symptoms that include sub-diagnostic threshold conditions can complement research of clinical diagnoses. Improved classification systems of psychopathology could investigate a weighted approach to symptoms, toward a more dimensional classification of mood disorders.

Keywords

Bipolar affective disorders; epidemiology; genetics; mood disorders; subtypes.

Background

Mood disorders are common in the general population[1,2] and lead to significant impairment in the individual, as well as direct and indirect costs to society.[3] The episodic nature and intra-individual symptom heterogeneity of these conditions can make diagnosis based on subjective symptom reports challenging in early phases of the disorder.[4] DSM-5[5] diagnostic criteria specify that a diagnosis of bipolar disorder requires a distinct period of abnormally and persistently elevated, euphoric or irritable mood; the presence of a specified number of additional concurrent symptoms; and usually some degree of impairment. The additional symptoms in DSM-5 encompass inflated self-esteem or grandiosity, decreased need for sleep, increased talkativeness, racing thoughts, being easily distracted, increased goal-directed activity or psychomotor agitation and engagement in activities that hold the potential for painful consequences.[5] Bipolar disorder type 1 and type 2 are differentiated by the presence of mania (type 1) compared with hypomania (type 2), a condition less disruptive to life than mania.

Data-driven classifications

Epidemiological studies of bipolar spectrum disorders use questionnaires to ascertain symptoms, with various approaches proposed.[6–8] In the UK Biobank study,[9] questions based on DSM-IV criteria were used to assess presence and severity of symptoms,[10,11] and responses

can be used to determine potential current or past disease. Whereas both diagnostic and epidemiological classifications reflect common clinical understanding of mood disorders, the use of data-driven approaches could justify or optimise such classifications. Further explorations of mental health definitions could aid epidemiological studies to refine the cases into more homogenous groups for investigation.[12] Precise phenotypes (or disease endotypes) will be instrumental in the shift to precision medicine and patient-specific tailored treatments, based on a more data-centric approach to disease taxonomy, with various frameworks and solutions already proposed.[13–16]

Latent class analysis (LCA) is a model-based probabilistic method of identifying homogenous subgroups of individuals (termed 'classes') based on patterns in a set of categorical indicator variables. Previous studies have used this data-driven approach to identify subtypes of disease based on symptom data. A general population study of both manic/irritable and psychotic episode symptoms (N = 1846) identified five classes differentiated by the presence of irritability and psychotic experiences, as well as differential associations with sociodemographic and clinical characteristics.[17] Other clustering methods have also been used to inform data-driven distinctions between mood disorders, such as with longitudinal patterns of mood to identify individuals with bipolar disorder type 1.[18] Previous studies conducting LCA of symptoms have often lacked replication in external data-sets, or have been performed in small samples.

Aims

In this study, we conducted a data-driven exploratory analysis of latent structure in reported symptoms experienced during manic and/or irritable episodes. Our aims were two-fold: first, to identify latent classes with homogeneous clinical characteristics and functional outcomes that may have clinical or biological relevance independent of diagnostic categories; and second, to investigate the correspondence of latent classes with reported psychiatric diagnoses and genetic liability to those, to aid in refining commonly used epidemiological definitions of probable bipolar disorder.

Method

Study populations

UK Biobank

Study participants for the discovery analysis were drawn from the UK Biobank study. Briefly, the UK Biobank is a prospective cohort study of over 500 000 individuals across the UK. Participants were aged 40–69 years at recruitment in 2006–2010.[9] Genotype data was available for all UK Biobank participants.[19] Ethical approval was granted by the NHS North West Research Ethics Committee (reference 11/NW/0382). Written informed consent was obtained from all participants. In a follow-up, participants who had agreed to be recontacted were invited to complete an online mental health questionnaire (MHQ) in 2017, resulting in additional phenotypic data for 157 366 UK Biobank participants.[11]

Phenotype data

To characterise probable history of mood disorders, UK Biobank worked with experts in mental health epidemiology to devise a self-completed online questionnaire, as clinical interviews would have been unfeasible given scale of the study. Questions were taken from existing tools at the time of the questionnaire's creation, aiming to maximise international compatibility. Questions on mania/hypomania were adapted from the Structured Clinical Interview for DSM-IV Axis I Disorders, as described.[10] Participants answered questions on ever having experienced a manic/irritable episode, as described in Box 1.

Box 1   Mental health questionnaire questions

Period of manic/hypomanic mood (field #20501)
'Have you ever had a period of time when you were feeling so good, 'high', 'excited', or 'hyper' that other people thought you were not your normal self or you were so 'hyper' that you got into trouble?'

Period of irritable mood (field #20502)
'Have you ever had a period of time when you were so irritable that you found yourself shouting at people or starting fights or arguments?'

Participants that answered positively to either or both of the above questions were subsequently asked if they had experienced any of the eight following symptoms, during these episodes (field #20548), selecting all that might apply.
Please try to remember a period when you were in a 'high' or 'irritable' state and select all of the following that apply:

    I was more talkative than usual,
    I was more restless than usual,
    I needed less sleep than usual,
    My thoughts were racing,
    I was more creative or had more ideas than usual,
    I was easily distracted,
    I was more confident than usual,
    I was more active than usual.

Participants who answered positively to above fields were then asked about:

The longest duration of any such episode (field #20492): brief (<24 h), moderate (>24 h but <1 week) or extended (>1 week).

The disruptiveness of the episode (field #20493): not disruptive or disruptive (if participants reported that the episode required treatment, caused problems with work, relationships, finances, the law or other aspects of life).

Sociodemographic data on participant gender, age, smoking status, alcohol intake frequency, Townsend Deprivation Index (TDI; a measure of area-level deprivation as a proxy for socioeconomic status) and education level were extracted from participant responses to the baseline questionnaire (see Supplementary Appendix 1 available at https://doi.org/10.1192/bjp.2021.184).

In the MHQ, participants reported past diagnoses by a professional (field #20544) of several disorders, which were used to define seven broad diagnostic categories: attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), generalised anxiety disorder (GAD), depression, schizophrenia/psychosis, mania/bipolar disorder and personality disorder (see Supplementary Appendix 1). Neuroticism score was derived from responses to the baseline questionnaire (see Supplementary Appendix).[10]

Electronic health records linked to Hospital Episode Statistics, which contain hospital diagnoses recorded with the ICD-10 up until June 2020, were used to derive cases status for four broad disorder definitions: depression, schizophrenia/psychotic disorder, mania/bipolar disorder and dementia (see Supplementary Appendix 1).

Polygenic risk scores

Genetic data pre-processing and sample exclusions are described in Supplementary Appendix 1. We calculated polygenic risk scores (PRS) with PRSice v2 for Linux (PRSice-2; see https://www.prsice.info/),[20,21] with clumping ($r^2 < 0.1$ and 500 kb window) and a P-value threshold of 1 (all single nucleotide polymorphisms included) for all analyses. PRS were residualised for the first six genetic principal components and scaled to a mean of 0 and s.d. of 1. Summary results from genome-wide association studies of anxiety disorder,[22,23] ADHD,[23] ASD,[24] major depression,[25] bipolar disorder[26] and schizophrenia[27] were used, from studies that did not include UK Biobank data (Supplementary Table 1).

PROTECT – replication sample

We attempted replication of findings in the Platform for Research Online to Investigate the Genetics and Cognition in Aging (PROTECT) study. Briefly, the PROTECT study is a UK-based online participant registry with continuous, ongoing recruitment beginning in 2015, which tracks the cognitive health of older adults. Study participants must be aged >50 years, have no diagnosis of dementia and have access to a computer/internet. Beginning in 2015, 14 836 PROTECT study participants were invited to complete the same online MHQ as UK Biobank participants, as a pilot of the questionnaire before roll-out in the UK Biobank study. In subsequent PROTECT study enrolment between 2016 and 2019, 21 475 participants in total completed the MHQ. The PROTECT MHQ included the same questions as the UK Biobank study, on ever experiencing a period of manic and/or irritable mood (see Box 1). Ethical approval was granted by the London Bridge National Research Ethics Committee (reference 13/LO/1578). Written informed consent was obtained from all participants.

Phenotype data

Participant responses to the MHQ questions on ever having experienced a manic and/or irritable episode, along with the

corresponding response to symptoms and episode duration/disruptiveness, were extracted by the same derivation process as the UK Biobank study. Sociodemographic variables on gender, age, smoking, education level and alcohol consumption frequency were derived from responses to baseline questionnaires.

Genetic data

A subset of PROTECT study participants provided a saliva sample for genotyping. Genetic data pre-processing and sample exclusions are described in Supplementary Appendix 1. The total number of individuals with genetic data was 8272, after exclusions. PRS were calculated as for the UK Biobank study, residualised for the first six genetic principal components and rescaled.

Statistical analysis

LCA

LCA is a model-based method that estimates the distribution of an underlying unobserved categorical variable, hypothesised to explain the patterns of association between a set of discrete variables. The estimated categorical variable describes subgroups (termed 'classes') of individuals. The method estimates the posterior probabilities of an individual belonging to a particular latent class. LCA was run with the poLCA package[28] in R version 3.6.0 for Linux (R Project; see https://www.r-project.org/), which uses the maximum likelihood method. Models with increasing numbers of classes, beginning at 2 and up to 7, were compared for best fit by using the Bayesian information criterion and Akaike information criterion. The relative entropy (a measure of classification certainty ranging between 0 and 1) was used to assess separation between classes.[29] The eight binary symptom responses in participants reporting a manic and/or irritable episode were used as indicators in the LCA (responses to: more talkative, more restless, less sleep, racing thoughts, more creative, easily distracted, more confident, more active).

Multinomial logistic regression

Multinomial logistic regression was used to test for association between class membership as the outcome (based on most likely class membership probability), and sociodemographic variables, disorder diagnoses (self-reported or hospital) and PRS. Posterior probabilities of class membership were used as weights. Relative risk ratios were estimated for each class, compared with a reference class (the largest class). For categorical variables (education attainment, smoking and alcohol consumption), dummy coding was used for each level, with the reference level of each being all combined remaining levels.

Results

In the UK Biobank MHQ, 42 183 participants responded positively to the questions on a manic and/or irritable episode and completed the episode symptoms questions (Supplementary Table 2). Characteristics of this analytical subset and all MHQ respondents are shown in Table 1.

LCA

LCA was applied to the eight binary symptoms, as indicators, in the subset of participants reporting a manic and/or irritable episode ($N = 42\,183$). As the number of classes increased, Bayesian information criterion and Akaike information criterion both continuously decreased, with no minimum attained (Supplementary Table 3). Elbow/scree plots[30] (Supplementary Figure 1) indicated that either a four- or five-class model was the optimum model. Plotting the conditional probabilities for each indicator symptom showed that the additional class in the five-class model was distinct from the other four classes (Supplementary Figure 2). We therefore selected the five-class model as the optimum model.

Table 1 Comparison of sociodemographic characteristics in the subset of participants in the latent class analysis (subset) with responses to the stem question on ever having experienced a period of manic and/or irritable mood as well as the subsequent questions of symptoms, and the participants who completed the Mental Health Questionnaire (whole sample) in the UK Biobank

		LCA subset		Full MHQ sample	
		n	%	*n*	%
Total		42 183	–	15 1159	–
Gender	Female	24 402	58	85 557	57
	Male	17 779	42	65 597	43
	Not applicable/missing	2	0.005	5	0.003
Education	University degree	18 820	45	68 467	45
	A levels, NVQ, HNC or HND	14 411	34	49 677	33
	O levels or CSE	5901	14	21 169	14
	None	2654	6.3	10 424	6.9
	Not applicable/missing	397	0.94	1422	0.94
Alcohol consumption	Daily	9245	22	35 176	23
	Weekly	20 973	50	76 917	51
	Occasionally	9371	22	30 484	20
	Never	2551	6	8459	5.6
	Not applicable/missing	43	0.1	123	0.081
Smoking	Current	2877	6.8	7303	4.8
	Past	17 124	41	56 483	37
	Never	22 076	52	87 037	58
	Not applicable/missing	106	0.25	336	0.22
		Mean	s.d.	Mean	s.d.
TDI[a]		−1.4	3.0	−1.7	2.8
	Not applicable/missing	78		196	
Age[b]		53.9	7.8	55.9	7.7
	Not applicable/missing	2		5	

LCA, latent class analysis; MHQ, mental health questionnaire; NVQ, National Vocational Qualification; HNC, Higher National Certificate; HND, Higher National Diploma; CSE, Certificate of Secondary Education; TDI, Townsend Deprivation Index.
a. TDI values before rank normalisation (high values indicates increased deprivation).
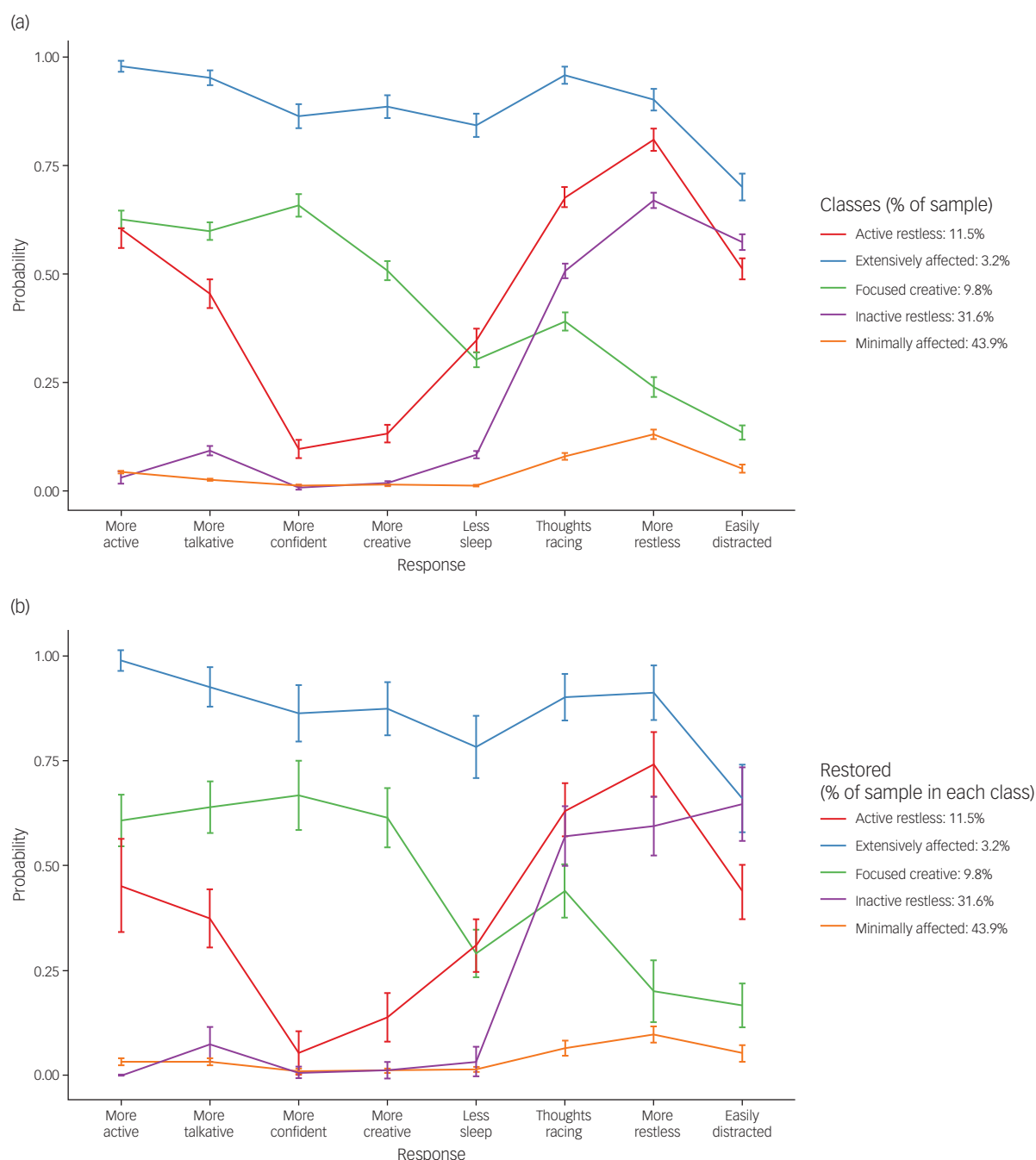b. Age at baseline.

Fig. 1 Conditional probabilities of each response (symptom) in (a) the UK Biobank optimum five-class latent model (*N* = 42 183) and (b) the Platform for Research Online to Investigate the Genetics and Cognition in Aging (PROTECT) replication study optimum five-class latent model (*N* = 4445).

The conditional probabilities of the eight indicator symptoms in each of the five latent classes are shown in Fig. 1(a). Individuals in the first class (3.2% of sample) had a high probability of reporting all symptoms and was therefore labelled the 'extensively affected' class. The second class (9.8%) was labelled 'focused creative', as individuals reported being more active, talkative, confident and creative. Individuals in the third class (11.5%) had high probabilities of being more active, talkative, restless, easily distracted and having racing thoughts. This class was labelled the 'active restless'. Individuals in the fourth class (31.6% of the sample) had a high probability of reporting racing thoughts, feeling more restless and being more easily distracted. This class was labelled the 'inactive restless'. The

fifth class (43.9%) had low probabilities of reporting all symptoms and was therefore labelled the 'minimally affected', and was used as the reference class in downstream analyses.

Distributions of responses to the original stem question of ever experiencing a period of manic and/or irritable mood by most likely class membership indicated that the inactive restless and minimally affected classes mostly comprised individuals reporting an irritable episode. The active restless class comprised individuals reporting an irritable episode and (to a lesser extent) both a manic and an irritable episode, whereas the focused creative class comprised individuals reporting an irritable, manic, or both a manic and irritable episode. The extensively affected class mostly comprised individuals
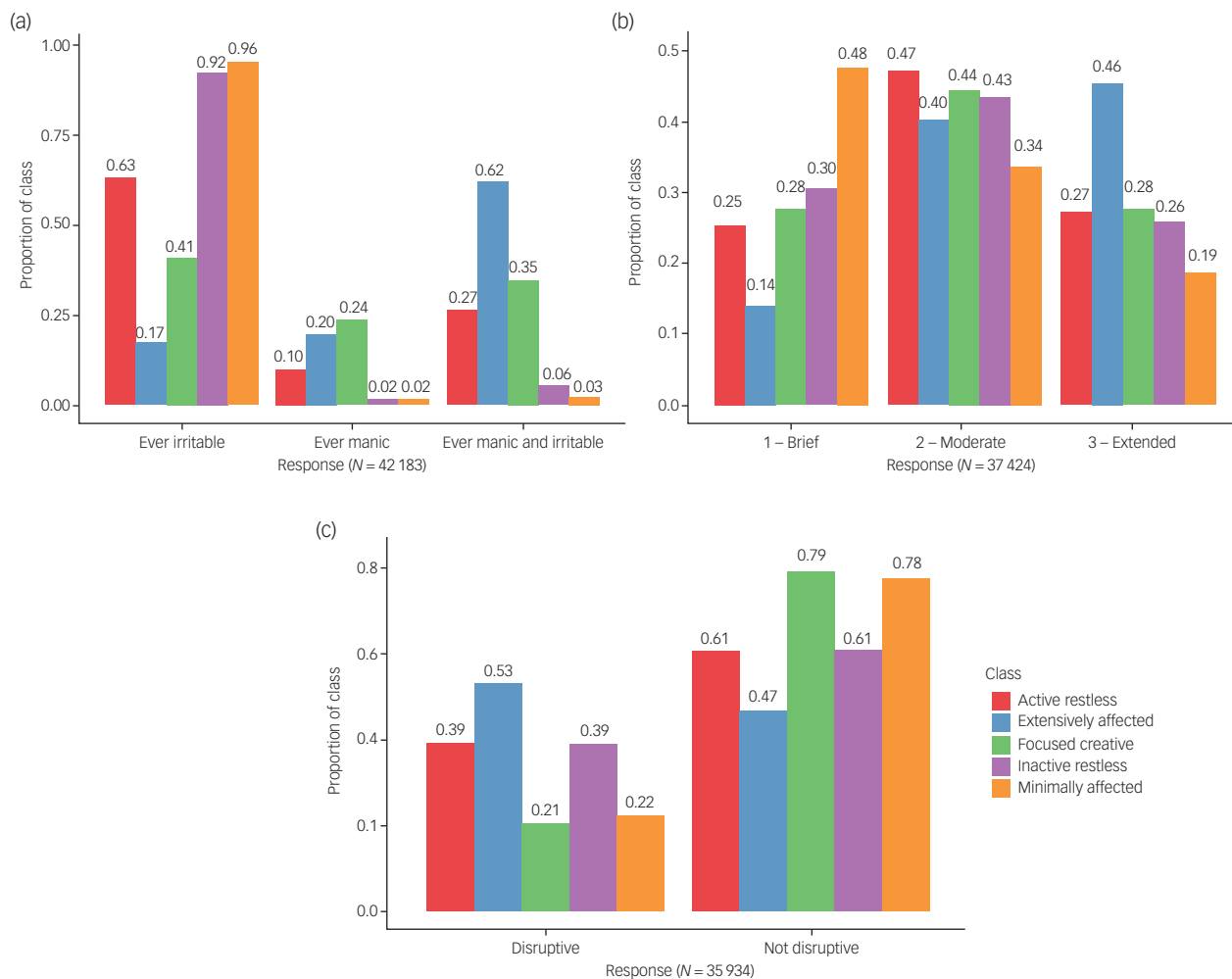
Fig. 2 Distributions of responses to (a) the stem question on ever experiencing a period of manic or irritable mood, (b) the subsequent question on episode duration in a subset of $n = 37\,424$ participants with available data and (c) the subsequent question on episode disruptiveness in a subset of $n = 35\,934$ participants with available data, by most likely class membership, in the optimum five-class model.

reporting both a manic and an irritable episode (Fig. 2(a), Supplementary Figure 3, Supplementary Table 4).

Associations with episode duration and disruptiveness

For responses to episode duration ($n = 37\,424$; brief, moderate or extended duration), individuals in the minimally affected class were more likely to report brief duration, whereas those in the extensively affected class mostly reported extended duration (Fig. 2(b)). Episode duration patterns did not substantially differ among the remaining three classes. Associations of episode duration with each class when using minimally affected as the reference largely reflected the observations from the most likely class membership (Supplementary Figure 5, Supplementary Tables 5 and 6).

Episode disruptiveness ($n = 35\,934$) showed a similar pattern to duration, with the highest proportion of reported disruption in the extensively affected class (53%) and lowest in the focused creative (21%) and minimally affected (22%) classes (Fig. 2(c), Supplementary Table 7). Individuals reporting disruptive episodes were more likely to be in the inactive restless and active restless classes, and far more likely to be in the extensively affected class (Supplementary Figure 7, Supplementary Table 8). Notably, levels of non-response to the questions on episode duration and disruptiveness were high ($n = 4759$ and $n = 6249$, respectively, Supplementary Figures 4 and 6).

Associations with sociodemographic characteristics

Associations with sociodemographic characteristics were investigated in a subset of $n = 41\,620$ individuals (Supplementary Tables 9–17, Supplementary Figures 8–16). Being male was associated with an increased risk of being in all other classes when compared with the minimally affected class, with a particularly high risk of being in the focused creative class. Higher educational attainment was associated with increased risk of being in the extensively affected and focused creative classes. For alcohol intake, individuals in the extensively affected and active restless classes were less likely to drink alcohol, whereas those in the focused creative class were more likely to drink daily. There was an increased risk of current smoking for the extensively affected class and a smaller increase for the remaining classes. For TDI score, there was an increased risk of being in the extensively affected class with increasing TDI score (increased deprivation), and smaller but significant increases in risk for the other classes, when compared with the minimally affected class.

Associations with self-reported diagnoses of psychiatric disorders

The self-reported diagnoses of six psychiatric disorders differed substantially between the latent classes ($N = 42\,183$). Over half of the

individuals (54.9%) did not report a diagnosis of any of the self-reported disorders studied: ADHD, GAD, ASD, mania/bipolar disorder, depression and schizophrenia/psychosis. Most individuals that did not report a diagnosis were members of the minimally affected (57%) or inactive restless (26%) classes. Among those that did report one or more diagnoses (Supplementary Figure 17), a diagnosis of either depression or GAD (or a combination of both) were the most numerous, and were mostly present in members of either the minimally affected or the inactive restless classes. Individuals with a diagnosis of mania/bipolar disorder, either alone or in combination with one or more of the remaining disorders, were mostly members of the extensively affected class. Diagnosis of any of the six disorders was associated with increased risk of being in the extensively affected class (Fig. 3(a)), with the highest increases in risk observed for mania/bipolar disorder and schizophrenia/psychosis. Diagnosis of depression and GAD was
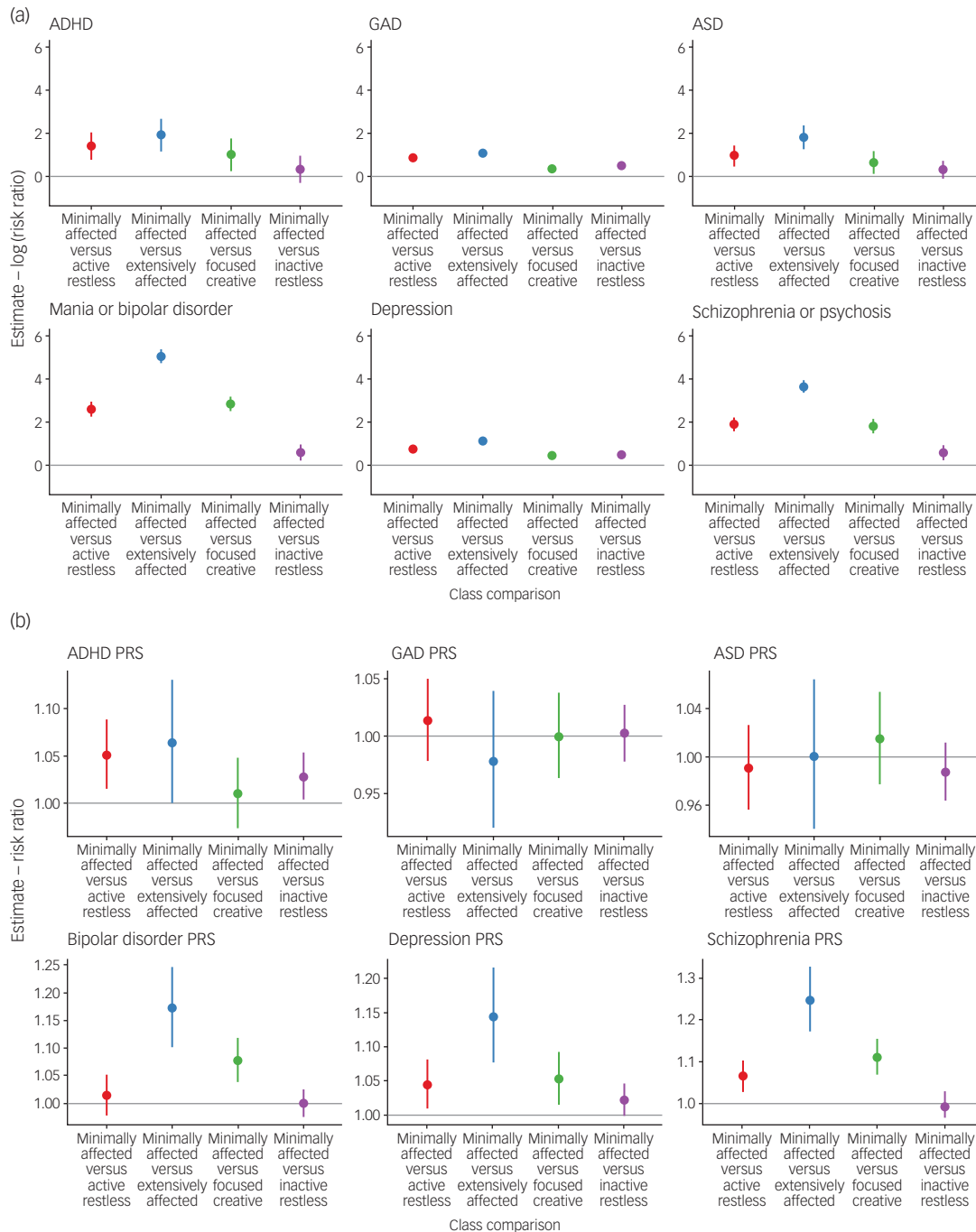

Fig. 3 (a) Associations of self-reported diagnoses of six disorders with most likely class membership, weighted for the probability of inclusion of an individual in that class. Effect estimates are presented as natural log risk ratio of inclusion in each class (relative to the reference class) for cases of each disorder. (b) Associations of PRS of six disorders with most likely class membership in a subset of $n = 33\,604$ with genetic data, weighted for the probability of inclusion of an individual in that class. Effect estimates are presented as risk ratios of inclusion in each class (relative to the reference class) per s.d. increase in standardised PRS for each disorder. The minimally affected class is used as the reference (comparison) class in all analyses. ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder; GAD, generalised anxiety disorder; PRS, polygenic risk score.

associated with increased risk of being in the inactive restless class, with weaker evidence for increased risk of being in this class for mania/bipolar disorder and schizophrenia/psychosis. Diagnosis of all six disorders was associated with increased risk of being in the focused creative and active restless classes, with the strongest associations for each class observed for mania/bipolar disorder (Supplementary Figure 18, Supplementary Tables 18–20).

Observed differences between classes when examining ICD-10 diagnoses of depression, mania/bipolar disorder and schizophrenia/psychotic disorder extracted from hospital records ($n = 36$ 258) largely corroborated findings of the analysis of self-reported diagnoses. For dementia diagnosis, we found little evidence for differences between classes. However, the number of cases of hospital diagnoses for all four disorders was low (Supplementary Figures 20 and 21, Supplementary Tables 21–23).

Associations of latent classes with self-reported diagnosis of personality disorder ($N = 42$ 183) indicated an increased risk of being in all classes when compared with the minimally affected class, with particularly large effects observed for the extensively affected class (Supplementary Figures 22 and 23, Supplementary Table 24). For derived neuroticism score, a 1 s.d. increase in score was associated with an increased risk of being in the extensively affected class, and with smaller but significant risk increases for all other classes, when compared with the minimally affected class (Supplementary Figure 24, Supplementary Table 25). We also explored overlap of each class with cases of probable bipolar disorder types 1 and 2, as defined by Davis et al[11] and Smith et al[10] (Supplementary Appendix 1, Results Section A, Supplementary Tables 24 and 25, Supplementary Figures 25 and 26).

Associations with PRS of psychiatric disorders

PRS of psychiatric traits discriminated between classes ($n = 33$ 604) (Fig. 3(b), Supplementary Tables 26 and 27). Schizophrenia PRS was associated with increased risk of being in the extensively affected, focused creative and active restless classes. For bipolar disorder PRS, there was an increased risk of being in the extensively affected and focused creative classes. Depression PRS conferred an increased risk of being in the extensively affected, focused creative and active restless classes. Results for ADHD were weaker, with an increased risk of being in the active restless and, to a lesser degree, inactive restless classes. Anxiety and ASD showed no significant increase in observed risk of being in any of the classes. These results contrast with the high proportion of GAD and ASD diagnoses reported by the extensively affected classes, but might also reflect lower power of the PRS for these disorders compared with the PRS of other disorders.

Replication

LCA

In the PROTECT replication cohort, there were $N = 4445$ participants with positive responses to the questions on ever experiencing a manic and/or irritable episode, approximately 10% of the sample size of the UK Biobank study. We observed some differences in characteristics between the studies, with a notably higher proportion of females in the PROTECT study than in the UK Biobank study (74% *v.* 58% in the analytical subsets) (Supplementary Tables 28 and 29).

Comparing latent class models with increasing numbers of classes indicated that a five-class model was again the optimum model, with an almost identical patterns of condition probabilities for the symptom indicators (Fig. 1(b), Supplementary Table 30). The size of some classes was notably different from the discovery cohort (31.6% *v.* 17% for inactive restless and 43.9% *v.* 56.9% for

minimally affected). Distributions of responses to the stem question of ever experiencing a period of manic and/or irritable mood were also similar to the discovery results. The inactive restless and minimally affected classes mostly comprised individuals reporting an irritable episode, whereas the extensively affected class mostly comprised both manic and irritable episodes. The focused creative and active restless classes were more mixed (Supplementary Figure 27, Supplementary Table 31).

Associations of latent classes in PROTECT

Similar associations to the discovery analyses were found between episode duration ($n = 3706$) and episode disruptiveness ($n = 3290$) with the five latent classes in the PROTECT study (Supplementary Figures 28–31, Supplementary Tables 32–35). Associations with sociodemographic characteristics ($n = 4411$) suggested similar distinctions between classes to the discovery analyses, although associations were often weaker and of smaller magnitude (Supplementary Figures 32–39, Supplementary Tables 36–39). For self-reported diagnoses of disorders ($n = 4421$), there were an adequate number of cases ($n > 20$) to analyse four disorders: depression, schizophrenia/psychosis, mania/bipolar disorder and GAD (Supplementary Figures 40–43). There was increased risk of being in all classes with a diagnosis of depression or GAD that mirrored the associations found in the discovery analysis. A diagnosis of schizophrenia/psychosis or mania/bipolar disorder led to an increased risk of being in the extensively affected class in particular (Supplementary Figures 44–47, Supplementary Tables 40 and 41). For PRS of six disorders ($n = 1494$), directions of effect were mostly consistent with the discovery cohort, but with confidence intervals that overlapped the null, except for an increased risk of being in the extensively affected class for bipolar disorder PRS (Supplementary Figure 48, Supplementary Table 42).

Discussion

Using a self-reported questionnaire based on diagnostic criteria for bipolar disorder, we have identified latent structure in participants reporting symptoms experienced during periods of manic and/or irritable mood. In both the main discovery cohort and the replication cohort, the participants were assigned to five latent classes. Class membership was associated with episode duration, episode disruptiveness, sociodemographic characteristics, diagnoses of psychiatric disorders and genetic risk of those disorders. These classes likely capture a broad range of disorders, as well as also sub-diagnostic threshold conditions and non-pathological experiences.

The extensively affected class comprises individuals who are the most markedly clinically affected, with particularly high prevalence of diagnoses of bipolar disorder and schizophrenia, as well as cases of depression, anxiety, ADHD and ASD. The inactive restless class comprises individuals with diagnoses of depression and anxiety, but fewer individuals with diagnoses of schizophrenia, bipolar disorder, ADHD or ASD. The active restless class comprises individuals with diagnoses of all disorders, to a lesser extent than the extensively affected class. The focused creative class comprises individuals with diagnoses of mostly bipolar disorder and schizophrenia, and to a lesser extent than the inactive restless class, anxiety and depression. Genetic analyses using PRS corroborate these findings, suggesting that the focused creative class has a higher genetic liability for bipolar disorder and schizophrenia, and the inactive restless class has a higher genetic liability for depression and ADHD. The minimally affected class may comprise individuals reporting normal variations in mood, with episodes of brief duration and low disruptiveness, with no increase in risk of disorder diagnosis

or genetic liability to any of the disorders. This minimally affected class may comprise individuals that experience symptoms that are not captured by the pre-defined questionnaire responses. As this was the largest class, our findings underline low specificity of the stem question in capturing clinically relevant periods of manic and/or irritable mood. Likewise, most participants who reported a manic and/or irritable episode, but not a mental health disorder, were in the minimally affected or inactive restless classes. The remainder of these individuals were members of the other three classes, indicating either underdiagnosis of mental health disorders, the presence of sub-diagnostic threshold symptoms or participant misreport of symptoms. Although we found little evidence of differences in dementia diagnosis between classes, mild cognitive impairment as a precursor to dementia diagnosis may lead to periods of irritable and/or manic mood. Longitudinal collections of cognitive measures in the UK Biobank study will enable future investigations of cognitive decline and class membership.

Contrasting dimensions of mood disorder symptoms were evident between classes. The active restless and inactive restless classes included disorganised, unproductive and unfocused characteristics, whereas the focused creative class included more creative characteristics, with higher education levels (similar to the extensively affected class) and lower levels of episode disruptiveness. Some psychiatric disorders have been suggested to share genetics with traits such as educational attainment[31] and creativity.[32,33] Participant responses to the questions in the MHQ are subjective and some participants, may perceive the symptoms they experience during episodes of manic/irritable mood less negatively than an external observer would.[34–36] However, this would not explain the more objective characteristic of higher educational attainment observed in the extensively affected and focused creative classes. Reported creative episodes and higher educational attainment in these two classes may precede onset and diagnosis of bipolar disorder, where the average age at onset for mood disorders is 29–43 (interquartile range 35–40) years of age.[2] Episodes of elevated mood experienced earlier in life may precede later-life bipolar disorder diagnosis and explain the observation. Further investigations into age at disorder onset and age at which episodes were experienced may aid in resolving these questions, with future follow-up questionnaires in the UK Biobank study extending the range of questions asked. Although results support a distinction between a less disruptive subtype of manic and/or irritable mood (the focused creative class) and more disruptive subtype(s) (e.g. the active restless and extensively affected classes), these classes cannot be mapped directly to bipolar disorder type 1 or 2 definitions. Instead, they suggest that the underlying symptoms can be used to group individuals into more homogenous classes, independently from a diagnosis of bipolar disorder. Future work should aim to further explore whether these homogenous groups can inform the debate on the distinction between bipolar disorder types 1 and 2,[37] or feed into new classification systems.

Symptom groupings in the LCA suggested some redundancy between possible responses in the questionnaire. Symptoms did not all contribute equally to class separation; for example, increased confidence and creativity appeared to differentiate the focused creative and extensively affected classes from the other classes, but did not separate out across classes. The five classes suggest that just four responses would suffice to distinguish the classes from each other, with symptoms forming the following groups: increased active/talkative, increased confident/creative, increased restless/thoughts racing/distracted and less sleep. These results may also inform research for future updates of the diagnostic classification systems. Rather than the current simple summation of number of symptoms present, a weighted approach to diagnostic criteria may be appropriate, constituting a step toward a more dimensional

classification of bipolar spectrum disorders. Although the five classes are categorical constructs, the underlying probabilities of individuals belonging to each class are on a continuous scale. The derived classes, as well as the more general latent structure reported among symptoms in our results, inform the ongoing development of novel classification systems, aiming to systematically evaluate the hierarchical taxonomy of disorders within psychopathology, and collate and integrate evidence generated across studies to date, such as the Hierarchical Taxonomy of Psychopathology (HiTOP)[16] and Research Domain Criteria (RDoC).[38] Future work could assess the merits of the current LCA approach against the use of continuous measurement instruments for symptom domains beyond manic/irritable episodes in bipolar spectrum disorders.

Our investigations have revealed differences in kind rather than just in degree between classes. Although a spectrum of increasing severity overlays the five classes, with the minimally affected class having the least severe presentation, we found higher numbers of cases of depression/anxiety in the inactive restless class and lower disruptiveness in the focused creative class, for example. Future work may further explore the effect of increasing psychopathology on class membership, particularly in relation to latent constructs such as the p-factor (general psychopathology factor).[39]

Strengths and limitations

There are a number of strengths to the present study. First, the use of a large, well-characterised cohort, the UK Biobank, ensures that results of this study will inform future mental health research in a well-powered, extensively studied and continually updated research resource. Second, the use of a model-based method enabled an agnostic bottom-up approach to defining latent subtypes that mitigates investigator bias of pre-defined criteria, and uses the data to inform selection of the number of optimum classes. Finally, the replication of the identified latent classes in an independent data-set, the PROTECT study, demonstrates robustness and replicability of the findings.

There are also several limitations that should be noted. First, the relative entropy of the optimum model in the UK Biobank and PROTECT studies was <0.7, indicating that classes may not be particularly homogenous, with some 'fuzziness' between classes. To account for this, we have weighted associations with the probability of belonging to each class in multinomial regressions. Entropy is usually not considered a model selection criterion and varies depending on the data under study.[29] Second, the study is limited by the scope of the questions that UK Biobank participants were asked on manic and/or irritable episodes experienced. Responses were dependent on the selection of multiple choice answers presented, and it is possible that other questions better characterise participant experiences, ultimately defining classes differently. However, since the DSM-5 uses similar symptom reports, the value of additional questions would be of limited clinical relevance at present. Third, given the use of two UK-based volunteer cohorts in restricted age groups (generally >50 years of age), generalisability beyond these populations is unknown. However, we would not expect age to substantially influence classes, because episode and symptom reports were lifetime retrospective. Finally, conclusions about associations with psychiatric diagnoses are limited by small numbers of individuals with hospital diagnoses, and in the replication data-set, low statistical power to fully replicate associations found in the discovery study.

We have used a data-driven approach, with replication in an external sample to derive latent classes differentiated by self-reported symptoms experienced during periods of manic and/or irritable mood that approximate the diagnostic criteria for bipolar

disorder. Our findings will inform future studies of mood disorders by guiding self-reported symptom data collection and interpretation, and research aimed at an improved characterisation of bipolar disorder in future classification systems of psychopathology.

Ryan Arathimos [ORCID], Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Chiara Fabbri** [ORCID], Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and Department of Biomedical and Neuromotor Sciences, University of Bologna, Italy; **Evangelos Vassos**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Katrina A. S. Davis** [ORCID], NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; and Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; **Oliver Pain**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Alexandra Gillett**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; **Jonathan R. I. Coleman** [ORCID], Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Ken Hanscombe**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Saskia Hagenaars**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; **Bradley Jermy**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; **Anne Corbett**, Faculty of Medicine, Department of Medicine, Imperial College London, UK; **Clive Ballard**, Medical School, College of Medicine and Health, University of Exeter, UK; **Dag Aarsland** [ORCID], Department of Old Age Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; and Centre for Age-Related Research, Stavanger University Hospital, Norway; **Byron Creese** [ORCID], Medical School, College of Medicine and Health, University of Exeter, UK; **Cathryn M. Lewis**, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK; NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, UK; and Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, UK

Correspondence: Ryan Arathimos. Email: ryan.arathimos@kcl.ac.uk

First received 14 Apr 2021, final revision 6 Nov 2021, accepted 16 Nov 2021

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1192/bjp.2021.184.

Data availability

Data are available from UK Biobank and PROTECT subject to standard access procedures (www.ukbiobank.ac.uk, www.protectstudy.org.uk/).

Acknowledgements

This paper represents independent research coordinated by the University of Exeter and King's College London. We gratefully acknowledge the participation of all National Institute for Health Research (NIHR) BioResource, the NIHR BioResource Centre Maudsley, Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London volunteers, and thank the BioResource staff for their help with volunteer recruitment. Further information can be found at https://www.maudsleybrc.nihr.ac.uk/facilities/bioresource/.

Author contributions

R.A. conducted the analyses and prepared the original draft of the paper. C.F., E.V., K.A.S.D., O.P., S.H., B.J., K.H. and J.R.I.C. were involved in data interpretation. A.G. was involved in methodology development and data interpretation. A.C., C.B., D.A. and B.C. were involved in data acquisition/collection and data curation. C.M.L. was involved in project conceptualisation and data interpretation. All authors reviewed, edited and approved the final draft of the paper.

Funding

Declaration of interest

None.

References

1 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey replication. *Arch Gen Psychiatry* 2005; **62**(6): 593–602.

2 Kessler RC, Angermeyer M, Anthony JC, de Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey initiative. *World Psychiatry* 2007; **6**(3): 168–76.

3 das Gupta R, Guest JF. Annual cost of bipolar disorder to UK society. *Br J Psychiatry* 2002; **180**: 227–33.

4 Phillips ML, Kupfer DJ. Bipolar disorder diagnosis: challenges and future directions. *Lancet* 2013; **381**(9878): 1663–71.

5 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (5th edn). American Psychiatric Association, 2013.

6 Hirschfeld RMA, Williams JBW, Spitzer RL, Calabrese JR, Flynn L, Keck J, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry* 2000; **157**(11): 1873–5.

7 Ghaemi SN, Miller CJ, Berv DA, Klugman J, Rosenquist KJ, Pies RW. Sensitivity and specificity of a new bipolar spectrum diagnostic scale. *J Affect Disord* 2005; **84**(2–3): 273–7.

8 Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. *J Affect Disord* 2005; **88**(2): 217–33.

9 Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**(3): e1001779.

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

11 Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. Mental health in UK Biobank – development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. *BJPsych Open* 2020; **6**(2): e18.

12 Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify psychiatric subtypes. *Trends Cogn Sci* 2019; **23**(7): 584–601.

13 Waszczuk MA, Eaton NR, Krueger RF, Shackman AJ, Waldman ID, Zald DH, et al. Redefining phenotypes to advance psychiatric genetics: implications from hierarchical taxonomy of psychopathology. *J Abnorm Psychol* 2020; **129**(2): 143–61.

14 van Os J, Delespaul P, Wigman J, Myin-Germeys I, Wichers M. Beyond DSM and ICD: introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry* 2013; **12**(2): 113–7.

15 Silbersweig D, Loscalzo J. Precision psychiatry meets network medicine network psychiatry. *JAMA Psychiatry* 2017; **74**(7): 665–6.

16 Kotov R, Waszczuk MA, Krueger RF, Forbes MK, Watson D, Clark LA, et al. The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J Abnorm Psychol* 2017; **126**(4): 454–77.

17 Pignon B, Peyre H, Noblins E, Rolland B, Thomas P, Vaiva G, et al. A latent class analysis of manic and psychotic symptoms. *J Nerv Ment Dis* 2020; **208**(9): 721–8.

18 Cochran AL, McInnis MG, Forger DB. Data-driven classification of bipolar I disorder from longitudinal course of mood. *Transl Psychiatry* 2016; **6**(10): e912.

19 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562** (7726): 203–9.

20 Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *GigaScience* 2019; **8**(7): giz082.

21 Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics* 2015; **31**(9): 1466-8.

22 Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry* 2016; **21**(10): 1391–9.

23 Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019; **51**(1): 63–75 .

24 Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* 2017; **8**: 21.

25 Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**(5): 668–681.

26 Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**(10): 977–83.

27 Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511** (7510): 421–7.

28 Linzer DA, Lewis JB. poLCA: an R package for polytomous variable latent class analysis. *J Stat Softw* 2011; **42**(10): 1–29.

29 Muthén BO. *Mplus Technical Appendices*. Muthén & Muthén, 1998–2004 (https://www.statmodel.com/download/techappen.pdf).

30 Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equ Modeling* 2007; **14**(4): 535–69.

31 Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DCM, Ritchie SJ, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112 151) and 24 GWAS consortia. *Mol Psychiatry* 2016; **21**(11): 1624–32.

32 Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci* 2015; **18**(7): 953–5.

33 Greenwood TA. Creativity and bipolar disorder: a shared genetic vulnerability. *Annu Rev Clin Psychol* 2020; **16**: 239–64.

34 Cooke MA, Peters ER, Kuipers E, Kumari V. Disease, deficit or denial? Models of poor insight in psychosis. *Acta Psychiatr Scand* 2005; **112**: 4–17.

35 Belvederi Murri M, Amore M. The multiple dimensions of insight in schizophrenia-spectrum disorders. *Schizophr Bull* 2019; **45**(2): 277–83.

36 Roe D, Kravetz S. Different ways of being aware of a psychiatric disability: a multifunctional narrative approach to insight into mental disorder. *J Nerv Ment Dis* 2003; **191**(7): 417–24.

37 Gitlin M, Malhi GS. The existential crisis of bipolar II disorder. *Int J Bipolar Disord* 2020; **8**: 5.

38 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**(7): 748–51.

39 Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* 2014; **2**(2): 119–37.

EXTRA CONTENT ONLINE · OPEN ACCESS · CC BY

Psychiatry in Literature

Georg Groddeck: the first psychoanalytical novel, *the Soul-Seeker*, and a musical mystery in Essex

Greg Wilkinson and Andrew Plant

Physician, pioneer of psychosomatic medicine, inspiration for the *Id*, after many rejections Groddeck succeeded in publishing *Der Seelensucher. Ein psychoanalytischer Roman* in 1921 with Internationale Psychoanalytische Verlag (IPV), at that time the most important publisher of psychoanalytic journals and books. Freud and Rank were shareholders and supported publication, whether or not, then and now, deemed pornographic and/or one of the funniest books in modern German literature. Compared by Freud to Cervantes and Rabelais, and by others to Swift and Balzac, reviewed favourably by Ferenczi, Freud's library catalogues two copies.

The Soul-seeker re-imagines psychoanalysis as Faustian satire and conduit for cultural liberation. The protagonist, August Mueller, personifies the *Id*. When he takes his recently widowed sister, Agathe, and her teenage daughter, Alwine, into his home, ineradicable bedbugs emerge, and infection in its origins and functions becomes the central metaphor. Maddened, August 'dies' and revives, un-infested, as Thomas Weltlein – a wise fool who leaves home and becomes involved in picaresque adventures through the strata of society. Rich sexual symbolism underpins motivation, thought and action in themes of art, childhood, crime, feminism, literature, medicine, philosophy, politics, religion and truth. Ultimately, our de-railed anti-hero perishes in a railway disaster. Ironically, his head cannot be found, and he is identified by a scar, its intimate location volunteered by Alwine, whose liaison with August invites interpretation.

Wit may be lost, and found, in translation. French and Italian versions are in currency, but the only one known in English appears to be by Christian Darnton (Baron von Schunck, 1905–1981), the communist Anglo-German composer and author, and his typescript remains repressed at The University of Essex in The Georg Groddeck Archive of Oscar Kollerstrom. Darnton experienced enduring adverse psychological effects of his traumatic childhood, described in his revealing memoir *No Name*. He consulted Groddeck, became a disciple, and translated *Der Seelensucher* with a view to its appearance in England. In March 1933, he wrote to the editorial board of IPV, which agreed to waive rights of translation if all other expenses were met – but there the publication trail ends, for now [EThOS ID: 490316].