medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Comparison of algorithm-based versus single-item phenotyping measures of depression and anxiety disorders in the GLAD Study cohort

Authors: Molly R. Davies[a,b], Joshua E. J. Buckman[c,d], Brett N. Adey[a,b], Chérie Armour[e], John R. Bradley[f], Susannah C. B. Curzons[a,b], Katrina A. S. Davis[a,b,g], Kimberley A. Goldsmith[a,b], Colette R. Hirsch[a,b,g], Matthew Hotopf[a,b,g], Christopher Hübel[a,b,h,i], Ian R. Jones[j], Gursharan Kalsi[a,b], Georgina Krebs[a,g], Yuhao Lee[a,b], Ian Marsh[a,b], Monika McAtarsney-Kovacs[a,b], Andrew M. McIntosh[k], Dina Monssen[a,b], Alicia J. Peel[a], Henry C. Rogers[a,b], Megan Skelton[a,b], Daniel J. Smith[l], Abigail ter Kuile[a,b], Katherine N. Thompson[a,b], David Veale[a,b,g], James T. R. Walters[d], Roland Zahn[a,b], Gerome Breen[a,b], Thalia C. Eley[a,b *]

[a]Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, Camberwell, London, UK

[b]National Institute for Health Research (NIHR) Biomedical Research Centre, South London and Maudsley Hospital, London, UK

[c]Centre for Outcomes Research and Effectiveness (CORE), Research Department of Clinical, Educational and Health Psychology, University College London, Gower Street, London WC1E 7HB, United Kingdom

[d]iCope – Camden and Islington Psychological Therapies Services, Camden & Islington NHS Foundation Trust, Finsbury Health Centre, Pine Street, London EC1R 0LP, United Kingdom

[e]Stress, Trauma & Related Conditions (STARC) research lab, School of Psychology, Queens University Belfast (QUB), Belfast, Northern Ireland, UK

[f]NIHR BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, UK

[g]South London and Maudsley NHS Foundation Trust, Denmark Hill, Camberwell, London, UK

[h]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[i]National Centre for Register-based Research, Aarhus Business and Social Sciences, Aarhus University, Aarhus, Denmark

[j]National Centre for Mental Health, Division of Psychiatry and Clinical Neuroscience, Cardiff University, Cardiff, UK

[k]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinurgh, UK

[l]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

*Corresponding author: Prof Thalia C. Eley, thalia.eley@kcl.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background: Research to understand the complex aetiology of depressive and anxiety disorders often requires large sample sizes, but this comes at a cost. Large-scale studies are typically unable to utilise "gold standard" phenotyping methods, instead relying on remote, self-report measures to ascertain phenotypes.

Aims: To assess the comparability of two commonly used phenotyping methods for depression and anxiety disorders.

Method: Participants from the Genetic Links to Anxiety and Depression (GLAD) Study (*N* = 37,419) completed an online questionnaire including detailed symptom reports. They received a lifetime algorithm-based diagnosis based on DSM-5 criteria for major depressive disorder (MDD), generalised anxiety disorder (GAD), specific phobia, social anxiety disorder, panic disorder, and agoraphobia. Any anxiety disorder included participants with at least one anxiety disorder. Participants also responded to single-item questions asking whether they had ever been diagnosed with these disorders by health professionals.

Results: Agreement for algorithm-based and single-item diagnoses was high for MDD and any anxiety disorder but low for the individual anxiety disorders. For GAD, many participants with a single-item diagnosis did not receive an algorithm-based diagnosis. In contrast, algorithm-based diagnoses of the other anxiety disorders were more common than the single-item diagnoses.

Conclusions: The two phenotyping methods were comparable for MDD and any anxiety disorder cases. However, frequencies of specific anxiety disorders varied depending on the method. Single-item diagnoses classified most participants as having GAD whereas algorithm-based diagnoses were more evenly distributed across the anxiety disorders. Future investigations of specific anxiety disorders should use algorithm-based or other robust phenotyping methods.

Abbreviations

CIDI (Composite International Diagnostic Interview)
CIDI-SF (Composite International Diagnostic Interview - short form)
SCID (Structured Clinical Interview for DSM-5)
MDD (major depressive disorder)
GAD (generalised anxiety disorder)
DSM-5 (Diagnostic Statistical Manual 5)
GLAD (Genetic Links to Anxiety and Depression)
EHR (electronic health records)
GP (general practitioner)

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Depression and anxiety disorders are common and debilitating, impacting approximately 30% of the population during their lifetime (1,2), and accounting for 10% of years lived with disability (3). This highlights the importance of understanding disorder-related risks and outcomes. In order to undertake research or treatment of these conditions, a vital step is identifying participants with or without the disorder of interest. The "gold standard" for phenotyping in psychiatric research is a structured or semi-structured diagnostic interview conducted in person (or over the phone) by a trained interviewer, such as the Composite International Diagnostic Interview (CIDI) (4) or Structured Clinical Interview for DSM-5 (SCID) (5). However, conducting in person interviews is time-consuming and costly. Due to the heterogeneous and complex aetiology of anxiety and depression, studies often require extremely large samples. This renders in-person interviews impractical and large-scale studies increasingly use online, self-report questionnaires to ascertain depression and anxiety disorder diagnostic status in participants.

There are two common methods to ascertain a diagnosis when using online questionnaires. *Algorithm-based diagnoses* involve a screening questionnaire which asks participants to self-report specific symptoms. The questionnaire responses are run through an algorithm based upon diagnostic criteria, such as the Diagnostic Statistical Manual (DSM-5; (6), to assess whether the participant qualifies for a diagnosis. This has been referred to as, variously, either strictly-defined, detailed, or symptom-based phenotyping (7–9). *Single-item diagnoses* take a contrasting approach and utilise a single question where participants are asked about the presence or absence of a clinical diagnosis from a health professional for a psychiatric disorder across their lifetime. They are also known as minimal, broad, or light-touch phenotyping (8,10). Both algorithm-based and single-item diagnostic methods are in widespread use in depression and anxiety research; however, it is unclear how they compare to one another. In this study, we compared algorithm-based and single-item lifetime diagnoses for major depressive disorder (MDD) and the five core anxiety disorders (generalised anxiety disorder (GAD), specific phobia, social anxiety disorder, panic disorder, and agoraphobia). Our aim was to assess agreement between these two phenotyping methods to determine to what extent they can be used interchangeably.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

Sample

The Genetic Links to Anxiety and Depression (GLAD) Study (https://gladstudy.org.uk) is an online research platform to recruit individuals with a lifetime experience of depression and/or anxiety for future research. The design and implementation of this study are described elsewhere (11). Recruitment is ongoing and this paper includes data from all participants that completed the survey as of May 19th, 2020 (N = 37,419). The average age of these participants was 38.1 years, 79.6% were female; the majority were white (94.5%), and a large proportion had a university degree (56.8%). Participants responded to an online, self-report questionnaire that included two methods for ascertaining likely depression and anxiety disorder diagnoses: algorithm-based and single-item.

Algorithm-based diagnoses

Algorithm-based diagnoses for MDD and GAD were evaluated using an adapted version of the short form Composite International Diagnostic Interview (CIDI-SF) (12) as used in UK Biobank (13). Similarly, items derived from the Diagnostic Statistical Manual (DSM-5) criteria assessed specific phobia, social anxiety disorder, panic disorder, and agoraphobia (14). Algorithms were developed to categorise participants as having a lifetime algorithm-based diagnosis for a disorder if their responses corresponded closely to DSM-5 criteria (see Appendix 1 in Supplementary Materials).

Single-item diagnoses

Single-item diagnoses were self-reported in response to the question: “Have you ever been diagnosed with one or more of the following mental health problems by a professional, even if you don't have it currently?” Participants were prompted to select all diagnoses that applied. Participants were categorised as having a single-item diagnosis if they selected the most comparable option to the relevant diagnosis (e.g., “Depression” for MDD). These single-item diagnoses reflect self-reports of a previous medically-provided diagnosis and were not validated against electronic health records (EHR). Phrasing for each of these items can be found in Appendix 2 in Supplementary Materials. We included the single-item of “panic attacks” as well as “panic disorder”, and separately compared both to algorithm-based panic disorder.

"Any anxiety" diagnosis

It is common in research to combine the anxiety disorder subtypes into a single category, arguing that the overlap between risk factors and outcomes is comparable (e.g., Purves et al[15]). We were interested in assessing agreement of algorithm-based and single-item diagnoses of “any anxiety” as well as that for the individual anxiety disorders. Algorithm-based “any anxiety disorder” was defined as participants with an algorithm-based diagnosis for at least one of the individual anxiety disorders (e.g., GAD, specific phobia, social anxiety disorder, panic disorder, or agoraphobia). Single-item diagnosis of “any anxiety disorder” included participants who self-reported receiving at least one anxiety disorder diagnosis from a health professional.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Analysis

We calculated the number of participants with zero, one, and two or more algorithm-based and single-item diagnoses. We also assessed the frequency of algorithm-based and single-item diagnoses for each disorder as percentages of the whole sample, excluding participants with missing data on one of the measures (e.g., a participant with single-item GAD but missing data for algorithm-based GAD was excluded from the GAD frequencies). Agreement and disagreement levels between these two phenotyping methods were assessed by calculating Cohen's kappa, sensitivity, and specificity. Sensitivity is the proportion of individuals with a disorder that the measure correctly classifies as having a diagnosis (proportion of true positives). In contrast, specificity is the proportion of individuals without a disorder that are correctly classified as not having a diagnosis (proportion of true negatives). Since we lacked a 'gold standard' reference in this sample, sensitivity and specificity analyses were conducted in both directions. All data cleaning and analyses were conducted in R version 3.5.3 (16).

Code availability

R scripts for the diagnostic algorithms and analyses included in this paper are available at https://github.com/mollyrdavies/GLAD-Diagnostic-algorithms.

Data availability

The data that support the findings of this study are available on request from the corresponding author, TCE. The data are not publicly available due to restrictions outlined in the study protocol and specified to participants during the consent process.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Frequencies

Frequency of single-item diagnoses were higher than algorithm-based diagnoses. As shown in Table 1, 35,399 (94.6%) participants reported a diagnosis of a major depressive or anxiety disorder on the single-item method (such a high proportion is expected since GLAD participants identified themselves as having had an anxiety and/or depressive diagnosis at some point in their lives), whereas 33,787 (89.8%) participants screened for at least one of the algorithm-based diagnoses. A higher proportion of participants (73.9%) reported two or more single-item diagnoses compared to two or more algorithm-based diagnoses (62.3%).

Table 1. Frequencies of algorithm-based and single-item diagnoses from the total sample.

		Algorithm-based diagnoses				Single-item diagnoses			
Subset	Total	0	1	2+	Mean	0	1	2+	Mean
All participants	37,419 (100%)	1,383* (5.0%)	10,443 (27.9%)	23,344 (62.4%)	2.16	1,463 (3.9%)	8,156 (21.8%)	27,243 (72.8%)	2.04
One or more single-item diagnosis	35,399 (94.6%)	1,185* (4.4%)	9,949 (28.1%)	20,475 (57.8%)	2.22				
One or more algorithm-based diagnosis	33,787 (90.3%)					1,079 (3.2%)	7,033 (20.8%)	25,673 (76.0%)	2.09

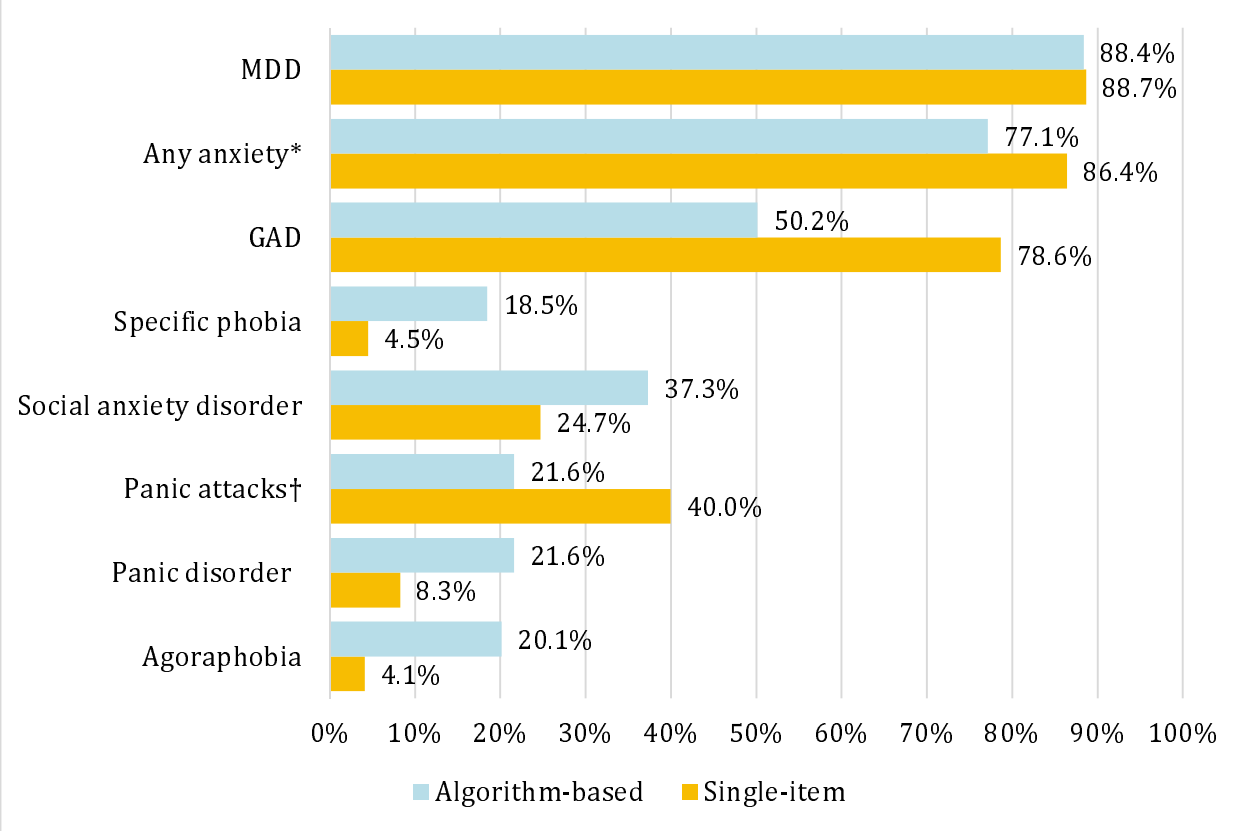
The table displays the number and percentage of subsets of GLAD participants (rows) with 0, 1, or 2+ algorithm-based or single-item diagnoses (columns). The mean number of algorithm-based and single-item diagnoses are reported. *For algorithm-based diagnoses, 9,754 participants had missing data for at least 1 disorder. Participants with at least 1 missing value for an algorithm-based diagnosis were excluded when calculating the percentage of participants with 0 algorithm-based diagnoses. Following this exclusion, 27,665 of total participants and 26,652 of participants with 1+ single-item diagnoses remained of which 1,383 (5.0%) and 1,185 (4.4%) respectively did not meet algorithm-based criteria for any disorder. Participants with missing data were included in the remaining frequencies for 1 or 2+ algorithm-based diagnoses.

Figure 1 displays the frequencies of algorithm-based and single-item diagnoses in the sample for each of the disorders. MDD had the highest frequency, which was consistent across phenotyping methods (88.4% algorithm-based, 88.7% single-item). The frequencies of the anxiety disorders varied widely depending on the measure. The majority of participants had a single-item diagnosis of GAD (78.6%), but the percentage of algorithm-based GAD diagnosis (50.2%) was approximately two-thirds of that, indicating a large discrepancy between the two methods. The remaining anxiety disorders had higher frequencies of algorithm-based than

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

single item diagnoses. For instance, the percentages of participants with algorithm-based specific phobia (18.5%), panic disorder (21.6%), and agoraphobia (20.1%) were more than double those of the respective single-item diagnoses. However, the proportion of algorithm-based panic disorder (21.6%) was only around half the frequency of single-item panic *attacks* (40.0%).

Figure 1. Frequencies of algorithm-based and single-item diagnoses of major depressive disorder, any anxiety, or an anxiety disorder in the GLAD sample.



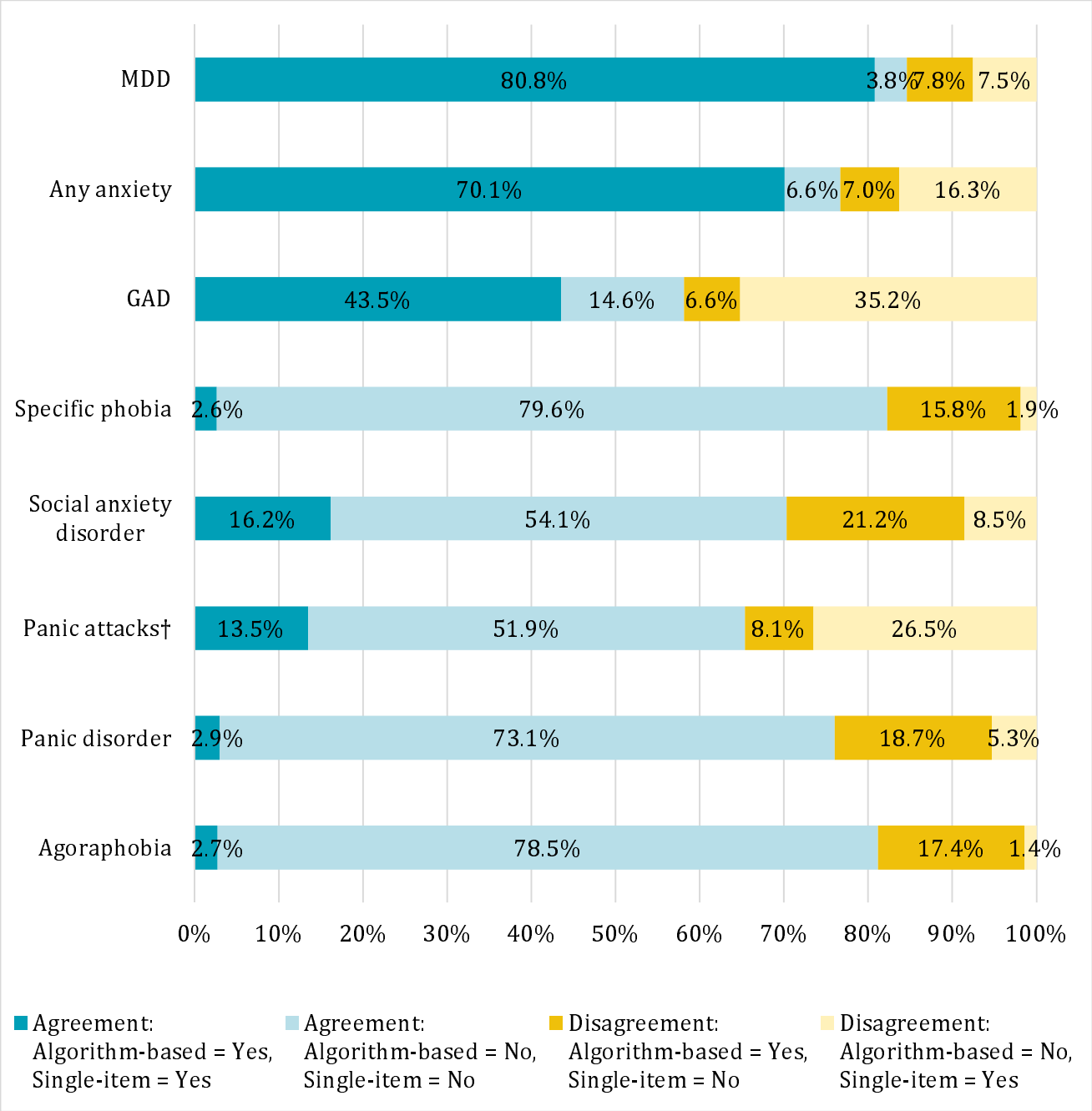
The bars represent the proportion (%) of GLAD participants (*N* = 37,419) with an algorithm-based (blue) or single-item diagnosis (yellow) for each disorder. *Any anxiety includes participants with at least one anxiety disorder (GAD, specific phobia, social anxiety disorder, panic disorder, and/or agoraphobia) on the indicated method (algorithm-based vs single-item). †For panic attacks, algorithm-based panic disorder is displayed and compared to single-item panic attacks.
Abbreviations: MDD, major depressive disorder; GAD, generalised anxiety disorder

Agreement

We examined the agreement between algorithm-based and single-item diagnoses. Figure 2 displays the agreement and disagreement for each disorder. The results in Figure 3 were also examined post-hoc by sex, but differences were minimal (see Appendix 3 in Supplementary Materials). Sensitivity, specificity, and Cohen's kappa are presented in Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2. All comparisons of agreement and disagreement on algorithm-based vs single-item clinical diagnoses.



Each bar displays the proportions (%) of the sample with agreement or disagreement between the two measures for each disorder. Agreements are represented in blue (dark blue = agreement on diagnosis, light blue = agreement on no diagnosis) while disagreements are in yellow (dark yellow = algorithm-based but no single-item diagnosis, light yellow = single-item but no algorithm-based diagnosis). †The panic attacks column displays the agreement between algorithm-based panic disorder and single-item panic attacks.
Abbreviations: MDD, major depressive disorder; GAD, generalised anxiety disorder

Table 2. Agreement between algorithm-based and single-item diagnoses.

			Algorithm-based (AB)			SI -> AB		AB -> SI		Cohen's
			Yes	No	Agreement (%)	Sensitivity	Specificity	Sensitivity	Specificity	Kappa
Major Depressive Disorder	Single-item (SI)	Yes	**28391**	2757	84.6	0.91	0.33	0.91	0.33	0.24
		No	2651	**1332**						
Any anxiety	Single-item (SI)	Yes	**21586**	5015	76.7	0.91	0.29	0.81	0.48	0.23
		No	2165	**2025**						
Generalised Anxiety Disorder	Single-item (SI)	Yes	**13914**	11260	58.2	0.87	0.28	0.55	0.69	0.16
		No	2124	**4682**						
Specific Phobia	Single-item (SI)	Yes	**877**	629	82.3	0.14	0.98	0.58	0.83	0.17
		No	5260	**26426**						
Social anxiety disorder	Single-item (SI)	Yes	**5681**	3004	70.3	0.43	0.86	0.65	0.72	0.32
		No	7436	**19032**						
Panic attacks†	Single-item (SI)	Yes	**3001**	5906	65.4	0.62	0.66	0.34	0.86	0.22
		No	1813	**11561**						
Panic Disorder	Single-item (SI)	Yes	**657**	1185	76.0	0.14	0.93	0.36	0.80	0.09
		No	4157	**16282**						
Agoraphobia	Single-item (SI)	Yes	**931**	480	81.2	0.13	0.98	0.66	0.82	0.17
		No	5969	**26889**						

Cross tabulations are presented for each disorder, with algorithm-based (yes/no) in columns and single-item (yes/no) in rows. Agreements between algorithm-based and single-item diagnoses are in bold. Sensitivity and specificity of single-item for algorithm-based (SI -> AB) and of algorithm-based for single-item (AB -> SI) are presented. †Analyses for single-item panic attacks were conducted with algorithm-based panic disorder.
Abbreviations: AB, algorithm-based; SI, single-item

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

MDD had the highest overall agreement (84.6%) between algorithm-based and single-item diagnoses, whereas GAD had the lowest (58.2%). However, Cohen's kappa values for all diagnoses were low (0.09-0.32), meaning that the reliability between these measures for all disorders is minimal at best (17).

Sensitivity (proportion of true positives) was high, but specificity (proportion of true negatives) was low of *single-item* MDD (0.91; 0.33), any anxiety (0.91; 0.29) and GAD (0.87; 0.28) for the respective algorithm-based measure. This indicates that these *single-item* diagnoses had high proportions of both true and false positives when compared to the algorithm-based measure.

Notably, sensitivity and specificity of *algorithm-based* MDD (0.91; 0.33) was the same as that found for single-item diagnoses, meaning that proportions of true positives and true negatives between these measures are comparable for these disorders, regardless of the direction of comparison. Sensitivity and specificity values indicated that *algorithm-based* any anxiety (0.81; 0.48) had high proportions of true and false positives for single-item any anxiety.

In contrast to the findings for MDD, GAD and any anxiety, sensitivity of *single-item* diagnoses for specific phobia, social anxiety disorder, panic disorder, and agoraphobia was low (0.13-0.43) while specificity was high (0.86-0.98). For these anxiety disorders, single-item diagnoses had low proportions of true positives, high proportions of false negatives, and high proportions of true negatives for the corresponding algorithm-based diagnoses. The sensitivity of *algorithm-based* diagnoses for single-item GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia was low to moderate (0.36-0.65) while specificity was moderate (0.69-0.83). This demonstrates that the algorithm-based anxiety subtypes predicted true positives of single-item diagnoses at approximately random chance (50%) but were moderately better at classifying true negatives.

The *single-item* measure of panic attacks had moderate sensitivity (0.62) and specificity (0.66) for algorithm-based panic disorder, indicating that classification of true positives and true negatives was slightly above random chance. Single-item panic attacks had a higher proportion of true and false positives than single-item panic disorder (0.14; 0.93) for algorithm-based panic disorder.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

Overview

In this study, we examined the agreement and disagreement between lifetime algorithm-based and single-item diagnoses of MDD, any anxiety, and the five core anxiety disorders (GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia). Analyses were conducted in participants who had self-defined as having lifetime experience of a depressive and/or anxiety disorder. We also assessed how single-item panic attacks compared to algorithm-based panic disorder, to determine whether agreement was better or worse than single-item panic disorder. Single-item diagnosis refers to the self-report of a diagnosis from a clinician, whereas algorithm-based diagnosis is based on participant responses to symptom questions that are then assessed against DSM-5 criteria for the disorder. Since the anxiety subtypes are sometimes grouped together in research (e.g., (15)), we included the "any anxiety" category to compare agreement for anxiety disorders as a group as well as individually.

Our results showed high agreement between algorithm-based and single-item diagnoses for MDD (84.6%). The lowest agreement between the two measures was for GAD (58.2%). Agreement for any anxiety (76.7%) and the other anxiety subtypes (specific phobia, social anxiety disorder, panic disorder, and agoraphobia) were higher (70.3 - 82.3%). Results from the sensitivity and specificity analyses demonstrated that single-item MDD, any anxiety, and GAD tended to over-diagnose compared to the respective algorithm-based diagnosis. Interestingly, algorithm-based MDD and any anxiety also had high proportions of false positives (low specificity) for the single-item measures. This suggests that single-item and algorithm-based measures for MDD and any anxiety have high disagreement overall on participants not meeting diagnostic criteria.

In contrast, our results suggested that single-item specific phobia, social phobia, panic disorder, and agoraphobia tended to under-diagnose when compared to the respective algorithm-based measure. Many participants with algorithm-based diagnoses of these anxiety subtypes did not report a single-item diagnosis of the same disorder. The majority of participants reported a single-item diagnosis of GAD rather than one of the other anxiety subtypes.

As expected, single-item panic attacks had a high proportion of false positives when compared to algorithm-based panic disorder. Panic attacks are a symptom that can manifest in isolation (18) and are not specific to panic disorder. However, the sensitivity of single-item panic attacks was higher than single-item panic disorder for the algorithm-based measure, indicating that this broader diagnosis captured a higher proportion of participants with algorithm-based panic disorder.

Implications

Our findings demonstrated that algorithm-based and single-item diagnoses for MDD and any anxiety are reasonably comparable and have particularly high agreement on participants with a diagnosis of MDD. These findings suggest that single-item MDD and any anxiety may be comparable to algorithm-based diagnoses for identifying cases of MDD but differ in the classification of those without a diagnosis. This is useful in the context of the efficacy of broad

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

phenotyping of MDD, with a diversity of opinions as to its value and utility in the field. Some studies reported that participants ascertained using single-item measures of diagnosis or even treatment-seeking have high genetic overlap with algorithm-based or clinically-ascertained MDD samples (19,20), suggesting comparability between the measures. Other researchers have argued that broad depression phenotyping shows the same genetic overlap with neuroticism and therefore is not specific to MDD (8). Algorithm-based MDD has also been found to have significantly higher heritability than single-item MDD, suggesting that utilising the single-item measure could decrease the power to detect genetic effects despite the increase in sample size (8,21). These reduced heritability estimates could be partially explained by the low sensitivity of single-item for algorithm-based MDD and any anxiety, as misclassification dilutes the power of case-control analyses to detect differences between the samples (22,23). Combining multiple broad phenotyping measures (e.g., single-item diagnoses, single-item help-seeking questions, and self-reported antidepressant usage) has been shown to reduce misclassification and increase heritability of MDD cases to equal or exceed heritability estimates of algorithm-based MDD in the UK Biobank (21).

However, our results indicate algorithm-based and single-item diagnoses for the anxiety subtypes (GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia) differ substantially in classifying positive diagnoses. Single-item methods categorised the majority of participants as having GAD, whereas algorithm-based measures show more even distribution across the subtypes. The lower percentage of single-item diagnoses of the anxiety disorders (aside from GAD) could be due to a lack of treatment-seeking or recognition. Many individuals with symptoms do not seek treatment for mental health or related problems (24,25) and those that do more commonly discuss their problems with a general practitioner (GP) rather than a mental health professional (24). However, research has shown that there is an under-recognition of anxiety disorders, particularly by GPs (26–29). GPs have limited amounts of time and resources and lack specialised training (30) to conduct comprehensive assessments of anxiety symptoms. It is therefore possible that GPs encountering distressed patients may identify symptoms as “anxiety” without specifying a disorder. In the GLAD Study, the phrasing of the single-item GAD question encapsulates general nerves, worry, or anxiety to account for this, but may be over-estimating the number of participants given a specific GAD diagnosis as a result.

These findings have important implications for research studies investigating disorder-specific risk factors or outcomes for anxiety subtypes. Although some factors are largely shared between major depressive and anxiety disorders (e.g., genetic factors), others show more specificity (e.g., environment (31,32), treatment approaches (33)). As such, genetic research studies focussed on expanding sample sizes may find that single-item measures are sufficient, since many of the genetic influences are shared between major depressive and anxiety disorders (32,34). However, single-item or broad phenotyping for anxiety disorders tends to categorise the majority of participants as having GAD. Therefore, in order to understand disorder-specific risk factors or investigate treatment approaches for the anxiety subtypes, an algorithm-based or more stringent assessment (e.g., SCID interview) would be required.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Limitations

The GLAD Study has been successful in recruiting a large number of participants to complete detailed phenotyping measures, which has enabled us to complete this thorough comparison of these two types of measures. However, as with any study, there are limitations. Eligibility criteria for the study included having either an algorithm-based or single-item diagnosis for depression, anxiety, or other related psychiatric disorder (e.g., bipolar disorder, obsessive-compulsive disorder). By design, we therefore have a low representation of participants without MDD or any anxiety (see Figure 1 for frequencies). Specificity (proportion of true negatives) was low between the measures, suggesting that single-item and algorithm-based methods may differ in terms of who they categorise as *not* having a disorder. However, this sample may not be equipped to accurately estimate specificity due to the small number of participants without a diagnosis. Another study conducted in the UK Biobank, a cohort of older adults recruited from the general population, compared algorithm-based and single-item MDD and found much lower agreement between the measures (7). This difference in method agreement between the two studies may be due to the higher proportion of participants in the UK Biobank sample without a diagnosis. Notably, both the UK Biobank and the GLAD Study samples are disproportionately white and highly educated compared to the UK population. The GLAD Study sample is also disproportionately female. Exploration of measurement agreement in more representative samples, both in terms of population demographics and prevalence of psychopathology, would establish whether the high agreement we found between these measures is generalisable.

The algorithm-based and single-item diagnoses have not been compared to a 'gold standard' clinical interview. There has been minimal and conflicting evidence for the validity of the self-report CIDI-SF for MDD, which was utilised here to determine algorithm-based MDD. Some studies show comparable overlap between the self-report CIDI-SF with diagnostic interviews (35,36) while others do not (37). Other studies comparing single-item measures to clinical interviews have found moderate agreement of single-item MDD (38,39) but poor agreement for single-item anxiety disorders (24). As a result, we cannot make any conclusions about which diagnosis is more accurate from the analyses conducted here.

Further research is therefore required to validate these measures against 'gold standard' clinical interviews. Validation of these measures is key to ensuring that research findings are relevant to clinical practice. Nonetheless, it is worth noting that some researchers have argued that a 'gold standard' diagnosis does not exist. Even structured and semi-structured interviews may result in different classifications of diagnosis and estimates of population prevalence (40). Other validation methods for these measures are worth exploring, such as comparing the genetic overlap or by comparing against clinical outcome measures such as functional impairment or treatment response.

At this point we could not assess whether participants' self-report of a clinical diagnosis matched their clinical data nor which health professional provided the diagnosis (e.g., general practitioner or psychiatrist). Other studies which have utilised single-item diagnoses in the context of genetics have similarly done so without medical record validation (15,19,20).

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Furthermore, since individuals with depressive and anxiety disorders often do not present in clinic or go undiagnosed (24,25,41), reliance on health records alone is not a substitute for asking the participant. However, all GLAD Study participants have consented to providing medical record access, so this comparison could be conducted in our sample in the future.

Conclusion

Large-scale research projects that lack the resources to conduct 'gold standard' clinical interviewing commonly utilise algorithm-based and single-item phenotyping methods. We compared these two measures and found good comparability between algorithm-based and single-item MDD and "any anxiety" disorder for categorisation of participants with a diagnosis. However, in contrast, there was poor agreement between these two types of ratings for participants not having the relevant disorder. Of note, ascertainment of participants with diagnoses for the individual anxiety disorders was largely different depending on which phenotyping measure was applied. Our results suggest that single-item diagnoses may be sufficient for discovery of shared genetic effects, but investigation of disorder-specific factors or outcomes would require an algorithm-based or other strictly-defined measure. In designing future studies, including and combining multiple methods of ascertaining diagnostic status, such as single-item, algorithm-based, and EHR data, may yield more robust phenotypes and increase power for analyses (21).

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Authorship contribution statement

G.B., T.C.E., J.R.B., C.R.H., M.H., I.R.J., N.K., A.M.M., D.J.S., D.V., and J.T.R.W. designed the GLAD Study.
M.R.D., B.N.A, C.A., S.C.B.C, C.H., G.K., I.M., M.M.K., Y.L., D.M., H.C.R., and K.N.T. carried out the data collection.
M.R.D. and T.C.E. conceived of the presented idea.
K.A.S.D. and K.A.G. advised on analyses to be included in the paper and assisted with interpretation.
Y.L., C.H, K.N.T., and M.R.D. cleaned and prepared the data. M.R.D. analysed the data, assisted by A.J.P., M.S., and A.t.K.
J.E.J.B., G.K., D.V., and R.Z. provided vital clinical input on the diagnostic algorithms and interpretation of results.
M.R.D. and T.C.E. wrote the paper.
G.B. and T.C.E. jointly supervised this work.
All authors contributed to interpretation of results. All authors provided critical feedback on manuscript drafts and approved the final version.

Funding

This work was supported by the National Institute of Health Research (NIHR) BioResource [RG94028, RG85445], NIHR Biomedical Research Centre [IS-BRC-1215-20018], HSC R&D Division, Public Health Agency [COM/5516/18], MRC Mental Health Data Pathfinder Award (MC_PC_17,217), and the National Centre for Mental Health funding through Health and Care Research Wales. Prof Eley and Dr Breen are part-funded by a program grant from the UK Medical Research Council (MR/M021475/1). Dr Buckman was supported by a Clinical Research Fellowship from the Wellcome Trust (201292/Z/16/Z). Dr. Goldsmith receives funding from NIHR, MRC, NIH, and the Juvenile Diabetes Research Foundation (JDRF). Dr Krebs is funded by a Clinical Research Training Fellowship from the Medical Research Council (MR/N001400/1).

Declaration of competing interest

Prof Breen has received honoraria, research or conference grants and consulting fees from Illumina, Otsuka, and COMPASS Pathfinder Ltd. Prof Hotopf is principal investigator of the RADAR-CNS consortium, an IMI public private partnership, and as such receives research funding from Janssen, UCB, Biogen, Lundbeck and MSD. Prof McIntosh has received research support from Eli Lilly, Janssen, and the Sackler Foundation, and has also received speaker fees from Illumina and Janssen. Prof Walters has received grant funding from Takeda for work unrelated to the GLAD Study. Dr Zahn is a private psychiatrist service provider and co-investigator on a Livanova-funded observational study. He has received honoraria for talks at medical symposia sponsored by Lundbeck as well as Janssen. He collaborates with EMOTRA, EMIS PLC and Alloc Modulo. The remaining authors have nothing to disclose.

Acknowledgements

We thank the GLAD Study and NIHR BioResource volunteers for their participation, and gratefully acknowledge the NIHR BioResource centres, NHS Trusts and staff for their contribution. We thank the National Institute for Health Research, NHS Blood and Transplant,

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and Health Data Research UK as part of the Digital Innovation Hub Programme. This study presents independent research funded by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Further information can be found at http://brc.slam.nhs.uk/about/core-facilities/bioresource. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, HSC R&D Division, Department of Health and Social Care.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Bibliography

1. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci. 2015 Sep;17(3):327–35.

2. 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 Jun;62(6):593–602.

3. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimate [Internet]. Geneva: World Health Organization; 2017 [cited 2019 Jan 19]. Available from: http://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf;jsessionid=6191AD7B3C8C385CC25248456EC9BFB7?sequence=1

4. World Health Organization. Composite International Diagnostic Interview. Geneva, Switzerland: World Health Organization; 1990.

5. First MB, Williams JBW, Karg RS, Spitzer RL. Structured clinical interview for DSM-5—Research version (SCID-5). Arlington, VA: American Psychiatric Association; 2015.

6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: Amer Psychiatric Pub; 2013.

7. Davis KAS, Cullen B, Adams M, Brailean A, Breen G, Coleman JRI, et al. Indicators of mental disorders in UK Biobank-A comparison of approaches. Int J Methods Psychiatr Res. 2019 Aug 8;28(3):e1796.

8. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet. 2020 Mar 30;52(4):437–47.

9. Nagel M. Changing perspectives: Towards detailed phenotyping in genetics. 2020 May 1;

10. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016 Aug 1;48(9):1031–6.

11. Davies MR, Kalsi G, Armour C, Jones IR, McIntosh AM, Smith DJ, et al. The Genetic Links to Anxiety and Depression (GLAD) Study: Online recruitment into the largest recontactable study of depression and anxiety. Behav Res Ther. 2019 Oct 24;123:103503.

12. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H-U. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). Int J Methods Psychiatr Res. 1998 Nov;7(4):171–85.

13. 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 Feb 6;6(2):e18.

14. Byrne EM, Kirk KM, Medland SE, McGrath JJ, Colodro-Conde L, Parker R, et al. Cohort profile: the Australian genetics of depression study. BMJ Open. 2020 May 26;10(5):e032580.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

15. Purves KL, Coleman JRI, Meier SM, Rayner C, Davis KAS, Cheesman R, et al. A major role for common genetic variation in anxiety disorders. Mol Psychiatry. 2019 Nov 20;

16. R Core Team. R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria. 2018 [cited 2019 Sep 20]. Available from: https://www.r-project.org/

17. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.

18. Kessler RC, Chiu WT, Jin R, Ruscio AM, Shear K, Walters EE. The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2006 Apr;63(4):415–24.

19. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019 Feb 4;22(3):343–52.

20. 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 Apr 26;50(5):668–81.

21. Glanville KP, Coleman JR, Howard DM, Pain O, Hanscombe KB, Jermy B, et al. Multiple measures of depression to enhance validity of Major Depressive Disorder in the UK Biobank. medRxiv. 2020 Sep 22;

22. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS ONE. 2013 Oct 11;8(10):e76295.

23. Schork A, Hougaard D, Nordentoft M, Mors O, Boerglum A, Bo Mortensen P, et al. Exploring contributors to variability in estimates of SNP-heritability and genetic correlations from the iPSYCH case-cohort and published meta-studies of major psychiatric disorders. BioRxiv. 2018 Dec 4;

24. McManus S, Bebbington P, Jenkins R, Brugha T. Health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. Leeds: NHS Digital; 2016 Jan.

25. Rayner C, Coleman JRI, Purves KL, Cheesman R, Hübel C, Gaspar H, et al. Genetic influences on treatment-seeking for common mental health problems in the UK biobank. Behav Res Ther. 2019 Jun 15;121:103413.

26. Tylee A, Walters P. Underrecognition of anxiety and mood disorders in primary care: why does the problem exist and what can be done? J Clin Psychiatry. 2007;68 Suppl 2:27–30.

27. Vermani M, Marcus M, Katzman MA. Rates of detection of mood and anxiety disorders in primary care: a descriptive, cross-sectional study. Prim Care Companion CNS Disord. 2011;13(2).

28. Arikian SR, Gorman JM. A review of the diagnosis, pharmacologic treatment, and economic aspects of anxiety disorders. Prim Care Companion J Clin Psychiatry. 2001

medRxiv preprint doi: https://doi.org/10.1101/2021.01.08.21249434; this version posted January 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Jun;3(3):110–7.

29. Fernández A, Rubio-Valera M, Bellón JA, Pinto-Meza A, Luciano JV, Mendive JM, et al. Recognition of anxiety disorders by the general practitioner: results from the DASMAP study. Gen Hosp Psychiatry. 2012 Jun;34(3):227–33.

30. Baird B, Charles A, Honeyman M, Maguire D, Das P. Understanding pressures in general practice. O'Neill K, editor. The King's Fund; 2016 May.

31. Waszczuk MA, Zavos HMS, Gregory AM, Eley TC. The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. JAMA Psychiatry. 2014 Aug;71(8):905–16.

32. Hettema JM, Prescott CA, Myers JM, Neale MC, Kendler KS. The structure of genetic and environmental risk factors for anxiety disorders in men and women. Arch Gen Psychiatry. 2005 Feb;62(2):182–9.

33. Clark DA, Beck AT. Cognitive Therapy of Anxiety Disorders: Science and Practice. Updated. New York: The Guilford Press; 2011.

34. Morneau-Vaillancourt G, Coleman JRI, Purves KL, Cheesman R, Rayner C, Breen G, et al. The genetic and environmental hierarchical structure of anxiety and depression in the UK Biobank. Depress Anxiety. 2020 Jun;37(6):512–20.

35. Levinson D, Potash J, Mostafavi S, Battle A, Zhu X, Weissman M. Brief Assessment Of Major Depression For Genetic Studies: Validation Of Cidi-Sf Screening With Scid Interviews. Eur Neuropsychopharmacol. 2017;27:S448.

36. Patten SB. Performance of the Composite International Diagnostic Interview Short Form for major depression in community and clinical samples. Chronic Dis Can. 1997;18(3):109–12.

37. Carlbring P, Forslin P, Ljungstrand P, Willebrand M, Strandlund C, Ekselius L, et al. Is the Internet-administered CIDI-SF Equivalent to a Clinician-administered SCID Interview? Cogn Behav Ther. 2002 Jan;31(4):183–9.

38. Stuart AL, Pasco JA, Jacka FN, Brennan SL, Berk M, Williams LJ. Comparison of self-report and structured clinical interview in the identification of depression. Compr Psychiatry. 2014 May;55(4):866–9.

39. Sanchez-Villegas A, Schlatter J, Ortuno F, Lahortiga F, Pla J, Benito S, et al. Validity of a self-reported diagnosis of depression among participants in a cohort study using the Structured Clinical Interview for DSM-IV (SCID-I). BMC Psychiatry. 2008 Jun 17;8:43.

40. Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychol Med. 1999 Sep;29(5):1013–20.

41. Kessler D, Bennewith O, Lewis G, Sharp D. Detection of depression and anxiety in primary care: follow up study. BMJ. 2002 Nov 2;325(7371):1016–7.