Journal of Anxiety Disorders 85 (2022) 102491

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

journal homepage: www.elsevier.com/locate/janxdis

Comparison of symptom-based versus self-reported diagnostic measures of anxiety and depression disorders in the GLAD and COPING cohorts

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], Helena L. Davies[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 Lin[a,b], Ian Marsh[a,b], Monika McAtarsney-Kovacs[a,b], Andrew M. McIntosh[k], Jessica Mundy[a,b], 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[j], 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, UK*
[d] *iCope – Camden and Islington Psychological Therapies Services, Camden & Islington NHS Foundation Trust, St Pancras Hospital, London, UK*
[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, Edinburgh, UK*
[l] *Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK*

ARTICLE INFO

Keywords:
Anxiety disorders
Measurement
Major depression
Self-report
Online surveys
Phobias

ABSTRACT

Background: Understanding and improving outcomes for people with anxiety or depression often requires large sample sizes. To increase participation and reduce costs, such research is typically unable to utilise "gold-standard" methods to ascertain diagnoses, instead relying on remote, self-report measures.
Aims: Assess the comparability of remote diagnostic methods for anxiety and depression disorders commonly used in research.
Method: Participants from the UK-based GLAD and COPING NBR cohorts (N = 58,400) completed an online questionnaire between 2018 and 2020. Responses to detailed symptom reports were compared to DSM-5 criteria to generate symptom-based diagnoses of major depressive disorder (MDD), generalised anxiety disorder (GAD), specific phobia, social anxiety disorder, panic disorder, and agoraphobia. Participants also self-reported any prior diagnoses from health professionals, termed self-reported diagnoses. "Any anxiety" included participants with at least one anxiety disorder. Agreement was assessed by calculating accuracy, Cohen's kappa, McNemar's chi-squared, sensitivity, and specificity.

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; NIHR, National Institute for Health Research; GLAD, Genetic Links to Anxiety and Depression; COPING, COVID-19 Psychiatry and Neurological Genetics; NBR, National Institute for Health Research BioResource; EHR, electronic health records; GP, general practitioner.

* Correspondence to: Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK.
E-mail address: thalia.eley@kcl.ac.uk (T.C. Eley).

https://doi.org/10.1016/j.janxdis.2021.102491
Received 5 February 2021; Received in revised form 13 October 2021; Accepted 20 October 2021
Available online 26 October 2021

Results: Agreement between diagnoses was moderate for MDD, any anxiety, and GAD, but varied by cohort. Agreement was slight to fair for the phobic disorders. Many participants with self-reported GAD did not receive a symptom-based diagnosis. In contrast, symptom-based diagnoses of the phobic disorders were more common than self-reported diagnoses.
Conclusions: Agreement for MDD, any anxiety, and GAD was higher for cases in the case-enriched GLAD cohort and for controls in the general population COPING NBR cohort. For anxiety disorders, self-reported diagnoses classified most participants as having GAD, whereas symptom-based diagnoses distributed participants more evenly across the anxiety disorders. Further validation against gold standard measures is required.

1. Introduction

Anxiety and depressive disorders are common and debilitating, impacting approximately 30% of the population during their lifetime (Bandelow & Michaelis, 2015; Kessler et al., 2005), and accounting for 10% of years lived with disability (World Health Organization, 2017). This highlights the importance of understanding disorder-related risk factors 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 ascertaining disorder diagnoses 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) (World Health Organization, 1990) or Structured Clinical Interview for DSM-5 (SCID) (First, Williams, Karg, & Spitzer, 2015). However, conducting interviews is time-consuming and costly. Due to the heterogeneous and complex aetiology of anxiety and depression, studies often require extremely large samples to reach sufficient statistical power. This renders diagnostic interviews impractical, and large-scale studies increasingly use online, self-report questionnaires to ascertain anxiety and depressive disorder diagnostic status of participants.

There are two common methods to ascertain a diagnosis when using online questionnaires. *Symptom-based diagnoses* involve a questionnaire which asks participants to self-report specific symptoms (Davis et al., 2019). The questionnaire responses are then compared to diagnostic criteria, such as the Diagnostic Statistical Manual (DSM-5 (American Psychiatric Association, 2013)), to assess whether the participant meets criteria for a diagnosis. This has also been referred to as strictly-defined or detailed diagnosis (Cai et al., 2020; Davis et al., 2019). *Self-reported diagnoses* take a contrasting approach and utilise a single question in which participants are asked whether they have received a clinical diagnosis from a health professional for a psychiatric disorder during their lifetime (Davis et al., 2019). This is also known as minimal, broad, or light-touch diagnosis (Cai et al., 2020; Hyde et al., 2016). Both symptom-based and self-reported diagnostic methods are in widespread use in anxiety and depression research; however, it is unclear whether they identify the same individuals.

One field which has focused heavily on ways to ascertain large sample sizes is that of psychiatric genetics. Anxiety and depression are heritable disorders. Heritability refers to the proportion of variance of a trait or disorder that is attributable to genetic factors, and decades of work have shown that approximately 20–30% of the variance of anxiety and depression can be attributed to genetics (Hettema, Neale, & Kendler, 2001; Kendler, Gatz, Gardner, & Pedersen, 2006; Shimada-Sugimoto, Otowa, & Hettema, 2015; Sullivan, Neale, & Kendler, 2000). Due to the complex nature of the genetics of anxiety and depressive disorders, samples in the hundreds of thousands have been required for adequate statistical power to detect significant associations with genetic variants. Psychiatric genetic studies have therefore used a variety of approaches to determine case and control status, often combining multiple methods in meta-analyses, and have assessed whether these measures represent the same constructs by investigating the genetic correlation between them. Genetic correlations indicate whether the same genetic variants are associated with different traits, thus lending evidence about their similarity (van Rheenen, Peyrot, Schork, Lee, & Wray, 2019).

Most of this work has been conducted on major depressive disorder (MDD). Participants ascertained using self-reported diagnoses have high genetic overlap with symptom-based or clinically-ascertained MDD samples (Howard et al., 2019; Wray et al., 2018), suggesting comparability between individuals ascertained with the two measures. However, symptom-based MDD has significantly higher heritability than self-reported MDD. Higher heritability means more power to detect significant genetic effects. The higher heritability of symptom-based MDD suggests that there are differences between symptom-based and self-reported diagnoses, and also implies that utilising the self-reported measure could decrease the power to detect genetic effects (Cai et al., 2020; Glanville et al., 2020). If the two methods are not comparable, then measure selection or meta-analyses across cohorts with different ascertainment methods may impact the detection of genetic, as well as other (e.g., demographic, environmental, social), risk factors and outcomes. However, if instead the two methods are comparable, not only does this support meta-analyses across datasets using these two approaches, but it also reduces the burden on future participants, researchers, and clinicians in ascertaining diagnoses. Understanding the agreement between these measures is an important goal with clear implications across research and clinical fields.

In this study, we compared symptom-based and self-reported lifetime diagnoses for 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 diagnostic methods to determine whether they can be used interchangeably in research.

2. Methods

2.1. Sample

Data were examined from the National Institute for Health Research (NIHR) BioResource cohort (N = 59,161).

This included 41,708 participants who had been recruited as part of the Genetic Links to Anxiety and Depression (GLAD) Study (https://gladstudy.org.uk). The GLAD Study is an online research platform to recruit individuals with a lifetime experience of anxiety and/or depression. Recruitment began in September 2018 and was conducted via traditional and social media campaigns or participating NHS sites.

The remaining 17,453 participants were NIHR BioResource members that had taken part in the COVID-19 Psychiatry and Neurological Genetics study (COPING NBR). This included members of the Irritable Bowel Disease cohort (IBD; N = 3,313) and general population cohorts (N = 14,140). There were several methods of initial recruitment to the NIHR BioResource, including through blood donation centres.

Both studies were conducted entirely online. Eligibility was limited to those aged 16 and over and who lived in the UK. Eligibility for the GLAD Study also required a lifetime experience of an anxiety or depressive disorder. Assessment occurred at a single stage in which all participants responded to online, self-report questionnaires that included two methods for ascertaining likely anxiety and depressive disorder diagnoses: symptom-based and self-reported. The analyses presented in this paper include data from all participants who completed

the GLAD or COPING survey before 10th December 2020. Additional details of the design and implementation of the GLAD Study are described elsewhere (Davies et al., 2019).

2.2. Measures

2.2.1. Symptom-based diagnoses

Symptom-based diagnoses were evaluated using the MDD, GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia modules from an adapted version of the short form Composite International Diagnostic Interview (CIDI-SF) (Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998), as used in the UK Biobank (Davis et al., 2020) and Australian Genetics of Depression study (Byrne et al., 2020). The CIDI-SF is based on the DSM-5 criteria for the disorders. Some validation studies of the online, self-administered version of the CIDI-SF for MDD have shown comparable agreement between symptom-based MDD with diagnostic interviews (Levinson et al., 2017; Patten, 1997). However, another study found low agreement between the online CIDI-SF and structured interviews for all disorders (MDD, GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia) (Carlbring et al., 2002). Algorithms were developed to categorise participants as having a lifetime symptom-based diagnosis for a disorder if their responses on the CIDI-SF corresponded closely to DSM-5 criteria (see Appendix 1 in Supplementary Materials).

2.2.2. Self-reported diagnoses

Self-reported diagnoses were assessed by 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 or indicate "None of the above". Participants who did not respond to the self-reported measure therefore had missing data for all self-reported diagnoses. Participants were categorised as having a self-reported diagnosis if they selected the most comparable option to the relevant diagnosis (e.g., "Depression" for MDD). Phrasing for each of these items can be found in Appendix 2 in Supplementary Materials. These self-reported diagnoses reflect self-reports of a previous medically-provided diagnosis and were not validated against electronic health records (EHR). Validation studies for self-reported diagnoses have found moderate agreement of self-reported MDD (Sanchez-Villegas et al., 2008; Stuart et al., 2014) but poor agreement for self-reported anxiety disorders (McManus, Bebbington, Jenkins, & Brugha, 2016) with structured interviews.

For the GLAD cohort, self-reported panic disorder was added partway through data collection and is only included for GLAD participants who completed the questionnaire after 5th November 2018. Participants who signed up before that date were excluded from all agreement analyses for panic disorder. We included self-reported "panic attacks" as well as "panic disorder", and separately compared both to symptom-based panic disorder. We are mindful that panic attacks are transdiagnostic and not specific to panic disorder. Research has shown that patients who have a panic attack are more likely to seek help from physical health professionals (e.g., in hospitals) than mental health services (Katerndahl & Realini, 1995; Wang et al., 2005). However, recognition and diagnosis of panic disorder from physical health professionals is low (Fleet et al., 1996; Lynch & Galbraith, 2003). Given the higher recognition of panic attacks compared to panic disorder, we were interested in comparing agreement between self-reported diagnoses of panic attacks and panic disorder with symptom-based panic disorder.

2.2.3. "Any anxiety" diagnosis

It is common in research to combine the anxiety disorder subtypes into a single category, given that the risk factors and outcomes overlap considerably between them (e.g., Purves et al., 2020). We were interested in assessing agreement of symptom-based and self-reported lifetime diagnoses of "any anxiety" alongside that of the individual anxiety disorders. Symptom-based "any anxiety disorder" was defined as participants with a symptom-based diagnosis for at least one of the individual anxiety disorders (i.e., GAD, specific phobia, social anxiety disorder, panic disorder, or agoraphobia). Self-reported diagnosis of "any anxiety disorder" included participants who self-reported receiving at least one anxiety disorder diagnosis from a health professional.

2.3. Analysis

We calculated the number of participants with zero, one, two, and three or more symptom-based and self-reported diagnoses. Participants with at least one missing value for a symptom-based diagnosis were included in the frequencies for one, two, or three or more symptom-based diagnoses. However, they were excluded when calculating the number of participants with zero symptom-based diagnoses. Self-reported panic disorder was added partway through GLAD data collection resulting in 14,858 GLAD participants with missing data on this item. These participants with missing data on self-reported panic disorder were included in all self-reported disorder frequencies. Participants with missing data on the remaining self-reported diagnoses were excluded.

We also assessed the frequency of symptom-based and self-reported diagnoses for each disorder as percentages of the whole sample, excluding participants with missing data on one of the measures for the disorder in question (e.g., a participant with self-reported GAD but missing data for symptom-based GAD was excluded from GAD frequencies for both measures).

Linear regression models were built to assess associations between demographic variables and missing data for symptom-based diagnoses, and logistic regressions were used to assess associations with self-reported diagnoses.

Agreement and disagreement levels between these two diagnostic methods were assessed by calculating: accuracy (the proportion [%] agreement), Cohen's kappa (Cohen, 1960), McNemar's within-subjects chi-squared test (McNemar, 1947), sensitivity, and specificity. Cohen's kappa calculated reliability between the two methods. Values range from zero to one, with higher values indicating greater reliability (Landis & Koch, 1977). McNemar's test assessed whether differences in the predictive accuracy of self-reported and symptom-based diagnoses were statistically significant ($\alpha < 0.05$). 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. We interpreted results as "agreed positives/negatives" and "disagreed positives/negatives".

The proportions of agreement and disagreement between these measures were also examined post-hoc by sex and compared using chi-squared analyses.

2.4. Code availability

All data cleaning and analyses were conducted using R version 3.5.3 (R Core Team, 2018), the tidyverse (Wickham et al., 2019), and caret (Kuhn, 2021) packages. The full code for the diagnostic algorithms and analyses included in this paper are available at https://github.com/mollyrdavies/GLAD-Diagnostic-algorithms.

2.5. 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.

3. Results

3.1. Sample characteristics

Participants with missing data for sex (N = 754 GLAD only) or age (N = 31 GLAD; N = 6 COPING NBR) were excluded from analyses. The remaining sample included 58,400 participants. Table 1 displays the sample descriptives by cohort. The average age of participants was 43 years, 73% were female, the majority self-defined as white (95%), and a large proportion had a university degree (54%). Characteristics between the cohorts were compared with *t*-test and chi-squared analyses. Given the large sample sizes, all characteristics were significantly different between the cohorts. The only differences that were clinically meaningful were age and sex, with the GLAD sample having a younger mean age and a higher proportion of female participants.

3.2. Frequencies

The frequency of the number of anxiety or major depressive disorder diagnoses was examined (Table 2). Since panic attacks are not a disorder, they were excluded from self-reported disorder frequencies. For symptom-based diagnoses, 21,779 participants (13,486 GLAD and 8,313 COPING NBR) had missing data for at least one disorder. Participants with zero self-reported or symptom-based diagnoses who have missing data for at least one disorder on the respective measure (excluding self-reported panic disorder, which was added partway through data collection) are displayed in Table 2 in the appropriate "NA" column.

Being male, self-identifying as Mixed or Asian/Asian British, and having a lower level of educational attainment were significantly associated with more missing data on symptom-based diagnoses. No characteristics were meaningfully associated with missing data on self-reported diagnoses. Full details of these analyses and summary of the results can be found in Appendix 3 in Supplementary Materials, along with a table of missing data by diagnosis and frequencies of the number of missing symptom-based diagnoses.

Frequency of diagnosis varied by cohort. As shown in Table 2, only 3% of GLAD participants did not receive a symptom-based diagnosis and 4% did not report a self-reported diagnosis for any disorder. However, for the COPING NBR sample these proportions were greater with 36% without a symptom-based diagnosis and 72% without a self-reported diagnosis. Of note, the proportion of COPING NBR participants without a symptom-based diagnosis was lower due to the high percentage (34%) in the missing data (NA) column, which included participants with no symptom-based diagnosis and missing data on the algorithm for at least one disorder. This difference in proportions between the cohorts is unsurprising, since GLAD participants were recruited for and therefore self-identified as having had an anxiety and/or depressive disorder diagnosis at some point in their lives whereas COPING NBR participants were recruited from the general population or for a physical health condition. Overall, 42,487 (73%) participants indicated a self-reported diagnosis of a major depressive or anxiety disorder, whereas 41,752 (71%) participants were identified as having at least one of the symptom-based diagnoses.

Fig. 1 displays the frequencies of symptom-based and self-reported diagnoses for the full sample and by cohort for each of the disorders. The bars for each diagnosis only include participants without missing data for either measure on the specified disorder. For instance, the proportion of participants with a MDD diagnosis was calculated as a percentage of those with no missing data for symptom-based and self-reported MDD.

MDD had the highest frequency for both cohorts, which was relatively consistent across diagnostic methods (GLAD: 88% symptom-based, 88% self-reported; COPING NBR: 29% symptom-based, 21% self-reported). The frequencies of the anxiety disorders varied widely depending on cohort and measure. The majority of GLAD participants had a self-reported diagnosis of GAD (78%) but the percentage with a symptom-based GAD diagnosis (59%) was approximately two-thirds of that value. In contrast, the percentage of COPING NBR participants with symptom-based (12%) and self-reported (13%) GAD was similar. The remaining anxiety disorders (the phobic disorders) had higher frequencies of symptom-based than self-reported diagnoses. For instance, the percentages of participants with symptom-based specific phobia (GLAD: 23%; COPING NBR: 5%), panic disorder (GLAD: 44%; COPING NBR: 5%), and agoraphobia (GLAD: 20%; COPING NBR: 2%) were more than double those of the respective self-reported diagnoses (GLAD: 4–8%; COPING NBR: 0.3–0.6%). The proportion of participants with symptom-based panic disorder (GLAD: 44%; COPING NBR: 5%) was more similar to self-reported panic *attacks* (GLAD: 40%; COPING NBR: 4%) than self-reported panic disorder (GLAD: 8%; COPING NBR: 0.6%).

3.3. Agreement

We examined the agreement between symptom-based and self-reported diagnoses. Fig. 2 displays the agreement and disagreement for each disorder. Accuracy, sensitivity, specificity, Cohen's kappa, and McNemar's test p-values are presented in Table 3.

Table 1
Sample characteristics.

Category	Descriptive	GLAD	% GLAD	COPING NBR	% COPING NBR	Full sample	% Full sample
Age	Mean	38***		54***		43	
	Standard deviation	14.47		14.42		16.12	
	Minimum	16		16		16	
	Maximum	93		99		99	
Sex	Male	8,335	20***	7,451	43***	15,786	27
	Female	32,618	80	9,996	57	42,614	73
Ethnicity	White, white European or Caucasian	38,087	94***	13,663	97***	51,750	95
	Mixed or multiple ethnic origins	1,033	3	160	1	1,193	2
	Asian or Asian British	573	1	184	1	757	1
	Black or Black British	207	0.5	56	0.4	263	0.5
	Arab	38	0.1	0	0	38	0.1
	Other	403	1	0	0	403	1
Highest education	University	21,213	55***	8,789	54***	30,002	54
	A-levels	9,078	23	3,177	19	12,255	22
	NVQ	3,453	9	1,836	11	5,289	10
	GCSE/CSE	5,079	13	2,589	16	7,668	14

Table 1 displays the sample characteristics for the Genetics Links to Anxiety and Depression (GLAD; N = 40,953) and COVID-19 Psychiatry and Neurological Genetics NIHR BioResource (COPING NBR; N = 17,447) cohorts, as well as the full sample (N = 58,400). Characteristics between the cohorts were compared using *t*-test and chi-squared analyses. Significant differences were indicated in the columns using the symbols below. Note that significant differences for ethnicity and highest education were not assessed for each factor level but represent overall differences.

*** $p < 0.001$.

Table 2
Frequencies of symptom-based and self-reported diagnoses from the full sample (N = 58,400) and by cohort.

		Symptom-based diagnoses[a]						Self-reported diagnoses[b]					
Subset	Total	0	1	2	3 +	NA	Mean	0	1	2	3 +	NA	Mean
Full sample	58,400	7,395 (12.66%)	13,461 (23.05%)	10,570 (18.1%)	17,721 (30.34%)	9,253 (15.84%)	1.78	14,268 (24.43%)	11,755 (20.13%)	20,032 (34.3%)	10,700 (18.32%)	1,645 (2.82%)	1.54
GLAD	40,953	1,158 (2.83%)	9,867 (24.09%)	9,296 (22.7%)	17,234 (42.08%)	3,398 (8.3%)	2.34	1,631 (3.98%)	8,725 (21.3%)	18,736 (45.75%)	10,357 (25.29%)	1,504 (3.67%)	2.04
COPING NBR	17,447	6,237 (35.75%)	3,594 (20.6%)	1,274 (7.3%)	487 (2.79%)	5,855 (33.56%)	0.45	12,637 (72.43%)	3,030 (17.37%)	1,296 (7.43%)	343 (1.97%)	141 (0.81%)	0.39
One or more symptom-based diagnosis	41,752							3,322 (7.96%)	9,435 (22.6%)	18,567 (44.47%)	10,223 (24.49%)	205 (0.49%)	1.94
One or more self-reported diagnosis	42,487	1,460 (3.44%)	11,138 (26.22%)	9,830 (23.14%)	17,257 (40.62%)	2,802 (6.59%)	2.32						

The table displays the number and percentage of the full sample, GLAD, and COPING NBR participants (rows) with 0, 1, 2, or 3 + symptom-based or self-reported diagnoses (columns). The mean number of symptom-based and self-reported diagnoses are reported. We also report the frequencies with 0, 1, 2, or 3 + symptom-based or self-reported diagnoses for participants with one or more diagnoses for the opposing measure (e.g., the number of participants with one or more symptom-based diagnoses who have 0, 1, 2, or 3 + self-reported diagnoses).

Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource.

[a] For symptom-based diagnoses, 21,779 participants (13,486 GLAD and 8,313 COPING NBR) had missing data for at least 1 disorder. Participants with at least 1 missing value for a symptom-based diagnosis were excluded when calculating the number of participants with 0 symptom-based diagnoses, and are instead included in the "NA" column under the "symptom-based diagnoses" header. Participants with missing data were included in the remaining frequencies for 1, 2, or 3 + symptom-based diagnoses.

[b] For self-reported diagnoses, the panic disorder option was added partway through data collection. Participants with missing data on any self-reported diagnosis except panic disorder were excluded from these frequencies, but those with missing data for panic disorder were included. The number of excluded participants is presented in the "NA" column under the "self-reported diagnoses" header. Panic attacks were not included in these figures.

In the full sample, we see a fairly similar pattern of effects for MDD, any anxiety, and GAD. The accuracy or proportion (%) agreement between symptom-based and self-reported diagnoses for these disorders was high (72–84%). Similarly, self-reported diagnoses had high sensitivity (0.83–0.87) and moderate specificity (0.63–0.75) for the respective symptom-based measure. This indicates that these *self-reported* diagnoses of MDD, any anxiety, and GAD had high proportions of agreed positives and slightly lower proportions of agreed negatives with the symptom-based measure. Notably, sensitivity and specificity of *symptom-based* MDD and any anxiety for self-reported diagnoses were similar, meaning that proportions of agreed positives and agreed negatives for these disorders were comparable between the measures, regardless of the direction of comparison. In contrast, symptom-based GAD had moderate sensitivity and high specificity for the self-reported measure. Self-reported GAD in the full sample therefore had higher sensitivity (i.e., proportion of agreed positives) for the symptom-based diagnosis than symptom-based GAD had for the self-reported diagnosis. These sensitivity results correspond with Fig. 2, which demonstrated that the largest proportion of disagreement for GAD (21%) were participants with a self-reported but not symptom-based diagnosis. Instead, MDD and any anxiety had equal proportions of the two types of disagreement, hence sensitivity and specificity results for these measures were similar in both directions. Despite the reasonable sensitivity and specificity values for MDD, any anxiety, and GAD in the full sample, Cohen's kappa indicated only moderate agreement for these measures (Landis & Koch, 1977).

When results for these three disorder categories are broken down by cohort, the proportion of agreement for MDD and any anxiety remained high for both GLAD and COPING NBR (81–84%), while the proportion of agreement for GAD was lower in GLAD (65%) than COPING NBR (87%). However, sensitivity and specificity results for all three disorders varied by cohort. In the GLAD cohort, *self-reported* MDD, any anxiety, and GAD had high sensitivity (0.86–0.91) and low specificity (0.33–0.35) for the symptom-based measures, whereas these self-reported diagnoses in the COPING NBR cohort had low to moderate sensitivity (0.44–0.56) and high specificity (0.92–0.94). For example, self-reported MDD had sensitivity of 0.91 and specificity of 0.33 in GLAD but sensitivity of 0.56 and specificity of 0.92 in COPING NBR. Self-reported MDD, any anxiety, and GAD therefore had high proportions of agreed positives in GLAD, and high proportions of agreed negatives in COPING NBR. Differences were observed in sensitivity and specificity results of the *symptom-based* diagnoses for the self-reported measures as well, with GLAD having higher sensitivity and lower specificity on each disorder than COPING NBR. For instance, symptom-based MDD had sensitivity of 0.91 and specificity of 0.33 in GLAD and sensitivity of 0.75 and specificity of 0.84 in COPING NBR for self-reported MDD.

Self-reported diagnoses of the phobic disorders (specific phobia, social anxiety disorder, panic disorder, and agoraphobia) showed a consistent pattern of results in the full sample and by cohort, displaying low sensitivity (0.08–0.43) and high specificity (0.86–1.00) for the symptom-based measures. Referring back to Fig. 2, the highest proportion of disagreement between the phobic disorder diagnoses was observed for participants with a symptom-based but not self-reported diagnosis. These results were more pronounced for the COPING NBR cohort, for which the self-reported diagnoses of the phobic disorders had sensitivity values below 0.17 and specificity around 1.00 for the symptom-based measure. Although the proportion of agreement for these disorders ranged between 61% (panic disorder in GLAD) and 98% (agoraphobia in COPING NBR), Cohen's kappa values indicated slight agreement for specific phobia, panic disorder, and agoraphobia diagnoses and fair agreement for social anxiety disorder in the full sample and by cohort.

The self-reported measure of panic attacks had higher sensitivity and lower specificity than self-reported panic disorder for symptom-based panic disorder in the full sample and by cohort. For example, in the full sample self-reported panic attacks had sensitivity of 0.61 and specificity of 0.86 while self-reported panic disorder had sensitivity of 0.15 and specificity of 0.98 for symptom-based panic disorder. Self-reported panic attacks therefore had a higher proportion of agreed positives but lower proportion of agreed negatives than self-reported panic disorder for the symptom-based measure, also observable in Fig. 2. Symptom-based panic disorder showed a comparable proportion of agreement with self-reported panic attacks and panic disorder, but Cohen's kappa indicated better agreement with self-reported panic attacks.

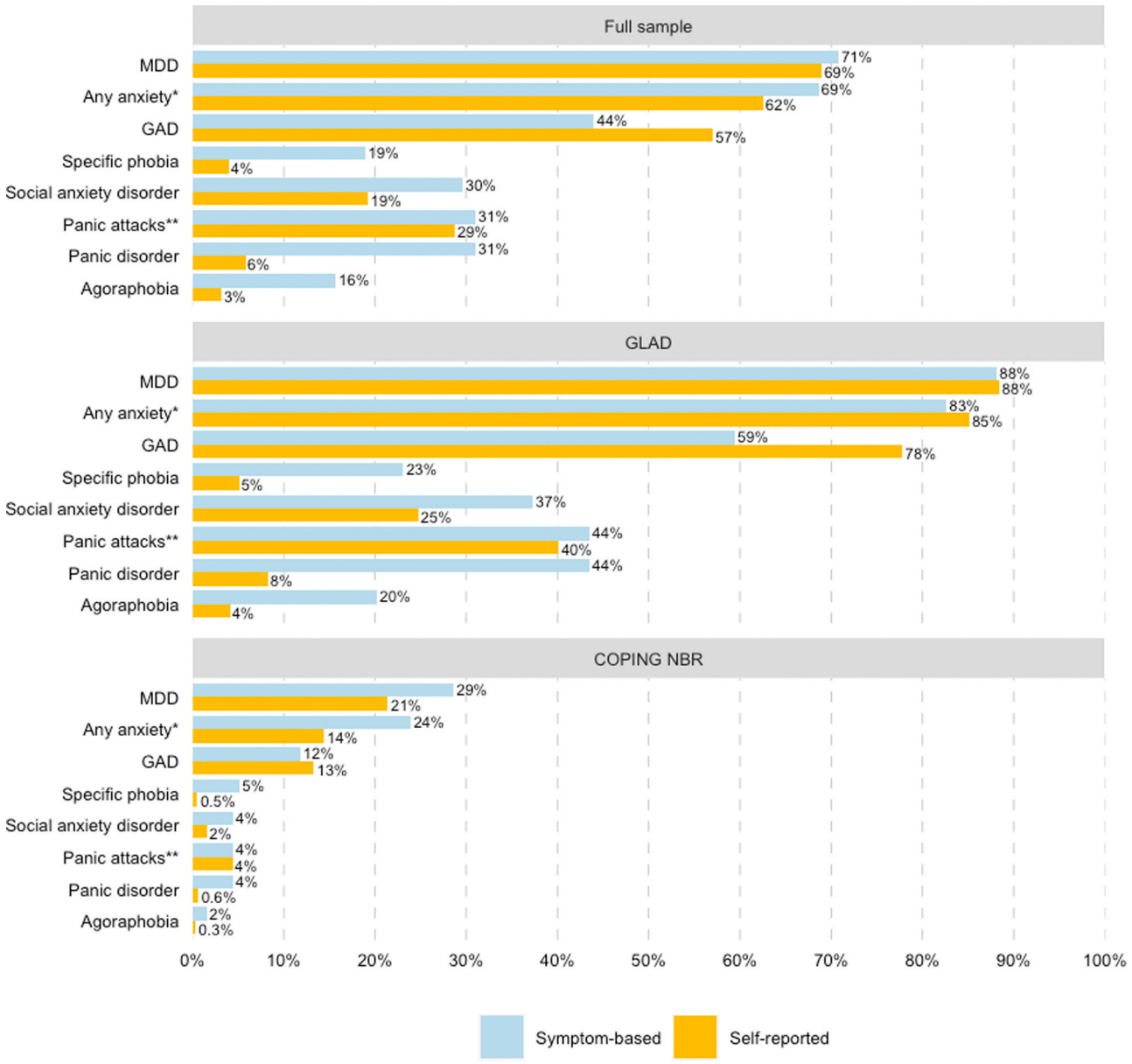

Fig. 1. Frequencies of symptom-based and self-reported diagnoses of major depressive disorder, any anxiety, or an anxiety disorder in the GLAD and COPING cohorts. The bars represent the proportion (%) of either the full sample (N = 58,400), GLAD (N = 40,953), or COPING NBR (N = 17,447) with a symptom-based (blue) or self-reported diagnosis (yellow) for each disorder. Proportions exclude participants with missing data on either measure for the specified diagnosis. *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 (symptom-based vs self-reported). * *For panic attacks, symptom-based panic disorder is displayed and compared to self-reported panic attacks. Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource; MDD, major depressive disorder; GAD, generalised anxiety disorder.

3.4. Sex differences in agreement

There was little variation by sex in proportions of agreement and disagreement between the measures (Appendix 4 in Supplementary Materials).

3.5. Alternate diagnoses

An unusually large proportion, relative to the other disorders, of the full sample with a self-reported diagnosis of GAD did not receive a symptom-based diagnosis (21%). We conducted a post-hoc analysis to explore whether these participants received any other symptom-based diagnosis. Fig. 3 displays the proportions (%) of participants with self-reported but not symptom-based GAD who had other symptom-based diagnoses, as well as the proportion without any symptom-based diagnosis. Proportions were calculated excluding participants with missing data for that symptom-based diagnosis. The proportion of participants with no symptom-based diagnosis excluded participants with missing data on any of the disorders.

The results showed that the largest proportion of participants with self-reported but not symptom-based GAD for both cohorts had symptom-based MDD (GLAD: 84%; COPING NBR: 59%). Of the anxiety disorders, symptom-based panic disorder was the most common for these participants (GLAD: 37%; COPING NBR: 15%). Over half of GLAD participants with self-reported but not symptom-based GAD had a different symptom-based anxiety disorder (58%), with only 8% having no symptom-based diagnosis. For COPING NBR, approximately one quarter of these participants had a different symptom-based anxiety

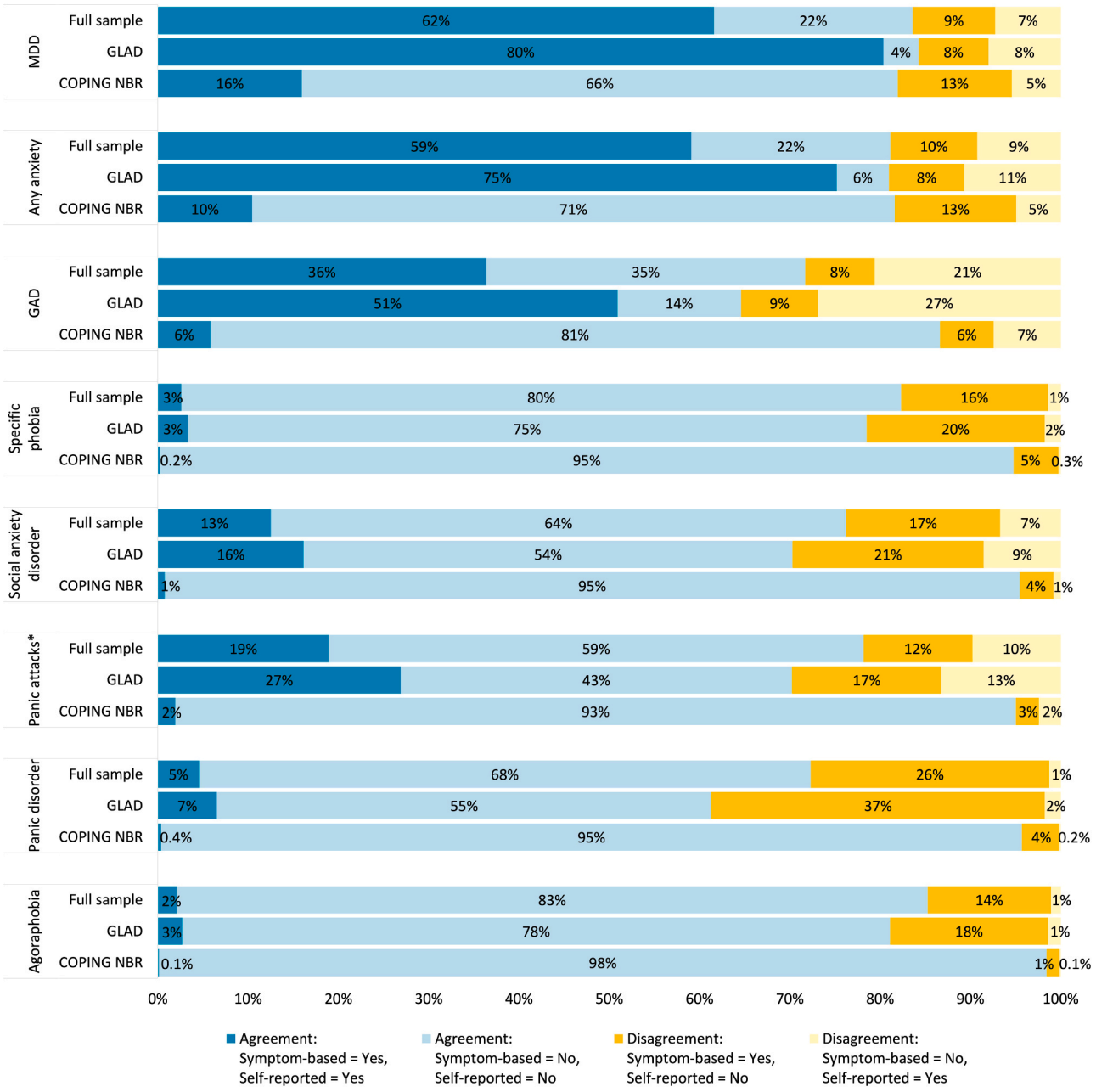

Fig. 2. All comparisons of agreement and disagreement on symptom-based vs self-reported diagnoses. Each bar displays the proportions (%) of the full sample (N = 58,400) and each cohort (GLAD: N = 40,954; COPING NBR: N = 17,447) with agreement or disagreement between the two measures for each disorder. Agreements are represented in blue (dark blue = agreement on diagnosis, light blue = agreement on no diagnosis) while disagreements are in yellow (dark yellow = symptom-based but not self-reported diagnosis, light yellow = self-reported but not symptom-based diagnosis). *The panic attacks column displays the agreement between symptom-based panic disorder and self-reported panic attacks. Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource; MDD, major depressive disorder; GAD, generalised anxiety disorder.

disorder (27%), but over one-third did not have any symptom-based diagnosis (34%).

The phobic disorders (specific phobia, social anxiety disorder, panic disorder, and agoraphobia) displayed relatively high proportions of participants with symptom-based but not self-reported diagnoses. For these disorders, we were interested in exploring whether these participants reported other self-reported diagnoses, different from the symptom-based diagnosis they received. We hypothesised that a large number of these participants would report a self-reported GAD diagnosis. Fig. 4 displays the proportions of participants with self-reported but not symptom-based diagnoses for specific phobia (N = 7,817), social anxiety disorder (N = 8,191), panic disorder (N = 9,215), and agoraphobia (N = 6,415) who indicated a self-reported diagnosis for another disorder.

Participants with a symptom-based but not self-reported diagnosis of the phobic disorders displayed a somewhat similar pattern of alternate diagnoses. Specifically, MDD and GAD were the most common alternate self-reported diagnoses for each phobic disorder, reported by over three-quarters of the GLAD cohort and between one-third and half of the COPING NBR cohort. Self-reported social anxiety disorder was also reported by over one-third of the GLAD cohort. For COPING NBR, in those with symptom-based but not self-reported agoraphobia, only social anxiety disorder was reported at a relatively higher rate. Alternate self-reported diagnoses of the other phobic disorders were rarely reported in

Table 3
Agreement between symptom-based and self-reported diagnoses for the full sample and by cohort.

				Symptom-based (SB)		Accuracy (%)	SR -> SB		SB -> SR			McNemar's test
Disorder	Cohort			Yes	No		Sensitivity	Specificity	Sensitivity	Specificity	Kappa	
MDD	Full sample	Self-reported (SR)	Yes	**32,533**	3,832	84 (83, 84)	0.87 (0.87, 0.87)	0.75 (0.74, 0.76)	0.89 (0.89, 0.90)	0.71 (0.70, 0.71)	0.61	$p < 0.001$
			No	4,841	**11,606**							
	GLAD	Self-reported (SR)	Yes	**30,077**	2,995	84 (84, 85)	0.91 (0.91, 0.92)	0.33 (0.31, 0.34)	0.91 (0.91, 0.91)	0.33 (0.32, 0.35)	0.24	$p \sim 0.21$
			No	2,898	**1,452**							
	COPING NBR	Self-reported (SR)	Yes	**2,456**	837	82 (81, 83)	0.56 (0.54, 0.57)	0.92 (0.92, 0.93)	0.75 (0.73, 0.76)	0.84 (0.83, 0.85)	0.52	$p < 0.001$
			No	1,943	**10,154**							
Any anxiety	Full sample	Self-reported (SR)	Yes	**25,304**	3,964	81 (81, 81)	0.86 (0.86, 0.86)	0.70 (0.70, 0.71)	0.86 (0.86, 0.87)	0.70 (0.69, 0.70)	0.56	$p \sim 0.08$
			No	4,122	**9,441**							
	GLAD	Self-reported (SR)	Yes	**24,189**	3,437	81 (81, 81)	0.90 (0.90, 0.90)	0.35 (0.34, 0.36)	0.88 (0.87, 0.88)	0.41 (0.39, 0.42)	0.26	$p < 0.001$
			No	2,690	**1,851**							
	COPING NBR	Self-reported (SR)	Yes	**1,115**	527	82 (81, 82)	0.44 (0.42, 0.46)	0.94 (0.93, 0.94)	0.68 (0.66, 0.70)	0.84 (0.83, 0.85)	0.42	$p < 0.001$
			No	1,432	**7,590**							
GAD	Full sample	Self-reported (SR)	Yes	**16,217**	9,197	72 (71, 72)	0.83 (0.82, 0.83)	0.63 (0.63, 0.64)	0.64 (0.63, 0.64)	0.82 (0.82, 0.83)	0.44	$p < 0.001$
			No	3,425	**15,756**							
	GLAD	Self-reported (SR)	Yes	**15,378**	8,128	65 (64, 65)	0.86 (0.85, 0.86)	0.34 (0.33, 0.35)	0.65 (0.65, 0.66)	0.62 (0.60, 0.63)	0.21	$p < 0.001$
			No	2,573	**4,133**							
	COPING NBR	Self-reported (SR)	Yes	**839**	1,069	87 (86, 87)	0.50 (0.47, 0.52)	0.92 (0.91, 0.92)	0.44 (0.42, 0.46)	0.93 (0.93, 0.94)	0.39	$p < 0.001$
			No	852	**11,623**							
Specific phobia	Full sample	Self-reported (SR)	Yes	**1,250**	691	82 (82, 83)	0.14 (0.13, 0.15)	0.98 (0.98, 0.98)	0.64 (0.62, 0.67)	0.83 (0.83, 0.83)	0.17	$p < 0.001$
			No	7,817	**38,331**							
	GLAD	Self-reported (SR)	Yes	**1,222**	660	78 (78, 79)	0.14 (0.14, 0.15)	0.98 (0.97, 0.98)	0.65 (0.63, 0.67)	0.79 (0.79, 0.80)	0.17	$p < 0.001$
			No	7,255	**27,639**							
	COPING NBR	Self-reported (SR)	Yes	**28**	31	95 (94, 95)	0.05 (0.03, 0.07)	1.00 (1.00, 1.00)	0.47 (0.34, 0.61)	0.95 (0.95, 0.95)	0.08	$p < 0.001$
			No	562	**10,692**							
Social anxiety disorder	Full sample	Self-reported (SR)	Yes	**6,011**	3,225	76 (76, 77)	0.42 (0.42, 0.43)	0.90 (0.90, 0.91)	0.65 (0.64, 0.66)	0.79 (0.78, 0.79)	0.36	$p < 0.001$
			No	8,191	**30,596**							
	GLAD	Self-reported (SR)	Yes	**5,923**	3,133	70 (70, 71)	0.43 (0.42, 0.44)	0.86 (0.86, 0.87)	0.65 (0.64, 0.66)	0.72 (0.71, 0.72)	0.32	$p < 0.001$
			No	7,767	**19,870**							
	COPING NBR	Self-reported (SR)	Yes	**88**	92	95 (95, 96)	0.17 (0.14, 0.21)	0.99 (0.99, 0.99)	0.49 (0.41, 0.56)	0.96 (0.96, 0.97)	0.24	$p < 0.001$
			No	424	**10,726**							
Panic attacks[a]	Full sample	Self-reported (SR)	Yes	**6,598**	3,408	78 (78, 79)	0.61 (0.60, 0.62)	0.86 (0.85, 0.86)	0.66 (0.65, 0.67)	0.83 (0.83, 0.84)	0.48	$p < 0.001$
			No	4,210	**20,665**							
	GLAD	Self-reported (SR)	Yes	**6,382**	3,136	70 (70, 71)	0.62 (0.61, 0.63)	0.77 (0.76, 0.77)	0.67 (0.66, 0.68)	0.72 (0.72, 0.73)	0.39	$p < 0.001$
			No	3,926	**10,276**							
	COPING NBR	Self-reported (SR)	Yes	**216**	272	95 (95, 95)	0.43 (0.39, 0.48)	0.97 (0.97, 0.98)	0.44 (0.40, 0.49)	0.97 (0.97, 0.98)	0.41	$p \sim 0.64$
			No	284	**10,389**							
Panic disorder	Full sample	Self-reported (SR)	Yes	**1,593**	447	72 (72, 73)	0.15 (0.14, 0.15)	0.98 (0.98, 0.98)	0.78 (0.76, 0.80)	0.72 (0.71, 0.72)	0.17	$p < 0.001$
			No	9,215	**23,626**							
	GLAD	Self-reported (SR)	Yes	**1,552**	424	61 (61, 62)	0.15 (0.14, 0.16)	0.97 (0.97, 0.97)	0.79 (0.77, 0.80)	0.60 (0.59, 0.60)	0.13	$p < 0.001$
			No	8,756	**12,988**							
	COPING NBR	Self-reported (SR)	Yes	**41**	23	96 (95, 96)	0.08 (0.06, 0.11)	1.00 (1.00, 1.00)	0.64 (0.51, 0.76)	0.96 (0.95, 0.96)	0.14	$p < 0.001$
			No	459	**10,638**							
Agoraphobia	Full sample	Self-reported (SR)	Yes	**980**	513	85 (85, 86)	0.13 (0.12, 0.14)	0.99 (0.99, 0.99)	0.66 (0.63, 0.68)	0.86 (0.86, 0.86)	0.18	$p < 0.001$
			No	6,415	**39,111**							
	GLAD	Self-reported (SR)	Yes	**965**	498	81 (81, 81)	0.13 (0.13, 0.14)	0.98 (0.98, 0.98)	0.66 (0.63, 0.68)	0.82 (0.81, 0.82)	0.17	$p < 0.001$
			No	6,253	**27,974**							
	COPING NBR	Self-reported (SR)	Yes	**15**	15	98 (98, 99)	0.08 (0.05, 0.14)	1.00 (1.00, 1.00)	0.50 (0.31, 0.69)	0.99 (0.98, 0.99)	0.14	$p < 0.001$
			No	162	**11,137**							

Cross tabulations are presented for each disorder for the full sample and by cohort, with symptom-based (yes/no) in columns and self-reported (yes/no) in rows. Agreements between symptom-based and self-reported diagnoses are in bold. Accuracy (%) and sensitivity and specificity of self-reported for symptom-based (SR -> SB) and of symptom-based for self-reported (SB -> SR), Cohen's kappa, and McNemar's test p-value results are presented. Accuracy, sensitivity, and specificity are reported with 95% confidence intervals in parentheses.

Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource; SB, symptom-based; SR, self-reported; MDD, major depressive disorder; GAD, generalised anxiety disorder.

[a] Analyses for self-reported panic attacks were conducted with symptom-based panic disorder.

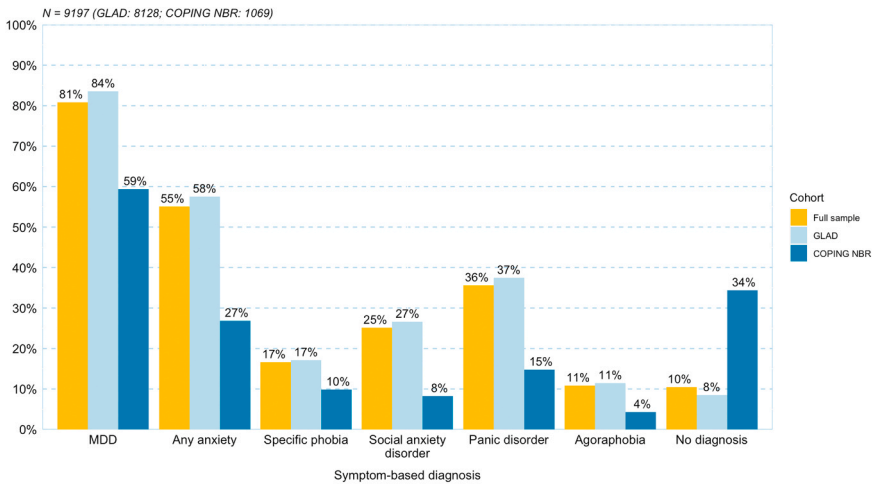

Fig. 3. Alternate symptom-based diagnoses for participants with self-reported but not symptom-based generalised anxiety disorder. Each bar displays the proportions (%) of participants with self-reported but not symptom-based GAD from the full sample (N = 9,197) and each cohort (GLAD: N = 8,128; COPING NBR: N = 1,069) with: symptom-based diagnoses for each of the other disorders, for any other symptom-based anxiety disorder, or without any symptom-based diagnosis (no diagnosis). The bars for the disorders are not exclusive as participants may have more than one symptom-based diagnosis. Proportions exclude participants with missing data for the indicated symptom-based diagnosis, and the proportion of "no diagnosis" excludes participants with missing data on any of the disorders. Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource; MDD, major depressive disorder; GAD, generalised anxiety disorder.

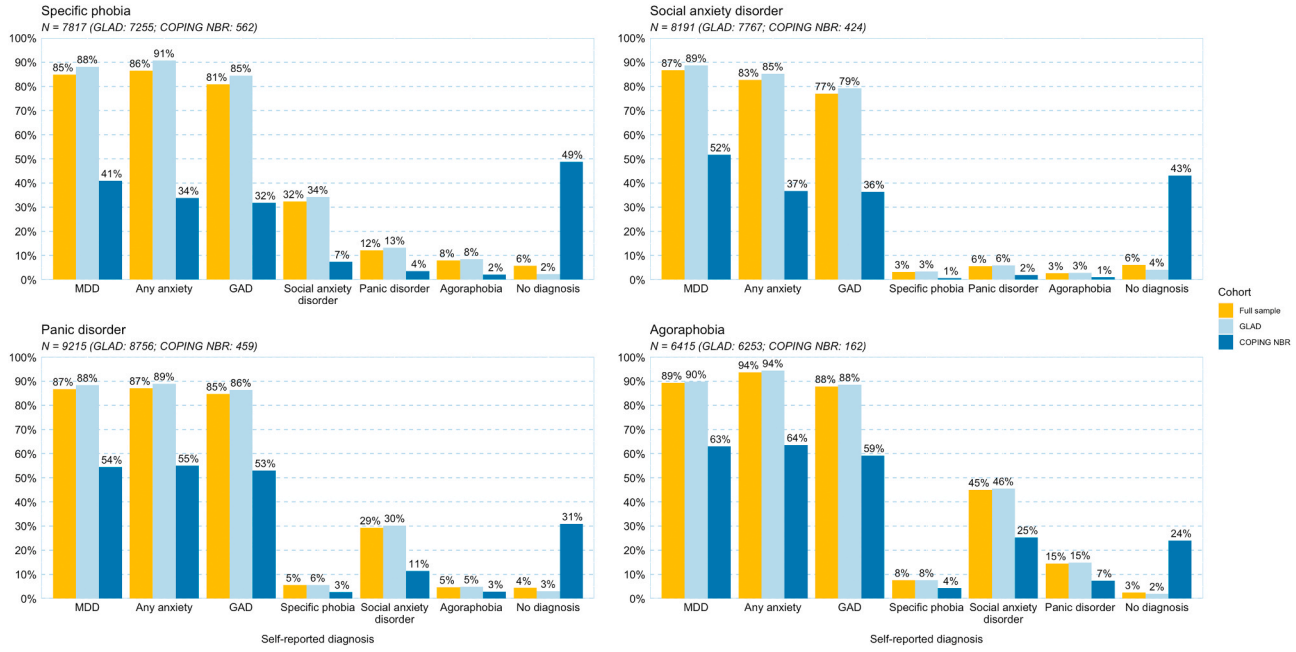

Fig. 4. Alternate self-reported diagnoses for participants with symptom-based but not self-reported specific phobia, social anxiety disorder, panic disorder, or agoraphobia. Each graph includes participants with a symptom-based but not self-reported diagnosis of specific phobia, social anxiety disorder, panic disorder, or agoraphobia. The titles of the plots indicate the disorder, and subtitles display the N of participants from the full sample and each cohort with a self-reported and no symptom-based diagnosis for the referenced disorder. The bars display the proportions (%) of these participants with self-reported diagnoses for each of the other disorders, for any other self-reported anxiety disorder, or without any self-reported diagnosis (no diagnosis). The bars for the disorders are not exclusive as participants may have more than one self-reported diagnosis. Proportions exclude participants with missing data for the respective self-reported diagnosis, and the proportion for "no diagnosis" excludes participants with missing data on any of the disorders. Abbreviations: GLAD, Genetics Links to Anxiety and Depression; COPING NBR, COVID-19 Psychiatry and Neurological Genetics NIHR BioResource; MDD, major depressive disorder; GAD, generalised anxiety disorder.

either cohort. Notably, between one-quarter and half of the COPING NBR cohort with a symptom-based but self-reported diagnosis of one of the phobic disorders did not have a self-reported diagnosis of any disorder, which was highly uncommon in the GLAD cohort (< 4%).

4. Discussion

4.1. Overview

In this study, we examined the agreement and disagreement between symptom-based and self-reported lifetime diagnoses of MDD, any anxiety, and the five core anxiety disorders: GAD, specific phobia, social anxiety disorder, panic disorder, and agoraphobia. These approaches are often utilised as the sole diagnostic methods in large-scale research studies and are frequently combined in meta-analyses despite limited evidence of comparability.

Symptom-based and self-reported diagnoses for MDD and any anxiety were reasonably comparable, demonstrating high accuracy (81–84%) and moderate agreement ($\kappa = 0.56$–0.61). Particularly high agreement was found for participants *with* a diagnosis (sensitivity 0.86–0.87). Although accuracy (72%) and Cohen's kappa ($\kappa = 0.44$) indicated moderate agreement for the GAD measures, a large proportion (21%) of the sample with self-reported GAD did not receive a symptom-based diagnosis, a finding not seen for MDD or any anxiety. This

parallels findings from 150,000 individuals recruited from the general population into the UK Biobank, which also found moderate agreement for MDD ($\kappa = 0.46$) but much lower agreement for GAD ($\kappa = 0.28$) (Davis et al., 2019). The self-reported measures for MDD, any anxiety, and GAD performed well at identifying *symptom-based cases* in the GLAD case cohort (sensitivity 0.86–0.90), but poorly in COPING NBR (sensitivity 0.44–0.56). In contrast, this approach performed well at identifying *symptom-based controls* in the COPING NBR general population cohort (specificity 0.92–0.94) yet was poor at identifying those without a diagnosis in GLAD (specificity 0.33–0.35). This suggests that the enrichment of cases (in GLAD) and controls (in COPING NBR) had a large impact on the performance of the two approaches.

In contrast, for the phobic disorders (specific phobia, social anxiety disorder, panic disorder, and agoraphobia) we found slight to fair agreement between measures ($\kappa = 0.17$–0.36). Sensitivity and specificity results were also notably different for self-reported diagnoses of the phobic disorders compared to MDD, any anxiety, and GAD. Self-reported phobic disorder displayed high agreement with the symptom-based measure on participants *without* a diagnosis (specificity 0.90–0.99), but poor agreement on participants *with* a diagnosis (sensitivity 0.13–0.42). Post-hoc analyses found that a relatively large proportion (12–26%) of the participants in our sample who received a symptom-based diagnosis of a phobic disorder did not self-report the same diagnosis; instead, many self-reported GAD or MDD.

Self-reported panic attacks had higher sensitivity and only slightly lower specificity than self-reported panic disorder for symptom-based panic disorder. These findings are contrary to what might be expected, since panic attacks are a symptom that can manifest in isolation (Kessler et al., 2006) and are not specific to panic disorder. However, self-reported panic attacks captured a higher proportion of participants with symptom-based panic disorder than self-reported panic disorder.

4.2. Limitations

A strength of the GLAD and COPING NBR studies is the successful recruitment of several thousand participants to complete detailed phenotyping measures. This enabled researchers to compare self-reported and symptom-based measures of depressive and anxiety disorders. However, as with any study, there are limitations that should be considered. Both cohorts are disproportionately female, white, and highly educated compared to the UK population. Exploration of measure agreement in more representative samples would establish the generalisablity of our findings.

As mentioned previously, the symptom-based and self-reported diagnoses have not been compared to a "gold standard" clinical interview in this study, and prior evidence is conflicting (Carlbring et al., 2002; Levinson et al., 2017; McManus et al., 2016; Patten, 1997; Sanchez-Villegas et al., 2008; Stuart et al., 2014). As a result, we cannot make any conclusions about which diagnostic method is more accurate from the analyses conducted here. Further research is therefore required to validate these measures against "gold standard" clinical interviews. Validation is key to ensuring that research findings are relevant to clinical practice. Nonetheless, it is worth noting that some researchers have argued that "gold standard" diagnoses do not exist; even structured and semi-structured interviews may result in different classifications of diagnosis and estimates of population prevalence (Brugha, Bebbington, & Jenkins, 1999). Other validation methods for these measures are worth exploring, such as investigating the genetic overlap with clinically-ascertained cohorts 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 [GP] or psychiatrist). In the context of genetics, studies that have utilised self-reported diagnoses have similarly done so without medical record validation (Howard et al., 2019; Purves et al., 2020; Wray et al., 2018). Furthermore, since individuals with anxiety and depressive disorders often do not present to a medical professional or receive a diagnosis (Kessler, Bennewith, Lewis, & Sharp, 2002; McManus et al., 2016; Rayner et al., 2019), reliance on health records alone is not a substitute for asking the participant. However, all GLAD and COPING NBR participants have consented to providing medical record access and an application for clinical data is underway, so this comparison could be conducted in future analyses.

4.3. Implications

We observed an asymmetry between cohorts in agreement results for MDD, any anxiety, and GAD, with agreement being stronger for cases in the case-enriched GLAD cohort and stronger for controls in the general population COPING NBR cohort. Consistent results across the cohorts were found for symptom-based and self-reported diagnoses of the phobic disorders (specific phobia, social anxiety disorder, panic disorder, and agoraphobia). The phobic disorder measures had high agreement for classification of participants *without* a diagnosis but differed substantially when classifying cases. Taken together, these findings suggest that studies on anxiety disorders applying self-reported diagnostic methods would tend to categorise the majority of participants as having GAD, whereas those utilising symptom-based measures would find more of a distribution across the anxiety subtypes.

These findings have important implications for large-scale studies investigating disorder-specific factors or outcomes, as ascertained diagnoses would differ depending on the selected measure. Although some factors are largely shared between anxiety and major depressive disorders (e.g., genetic factors), others show more specificity. For example, aspects of the environment show differential associations with anxiety and depression (Finlay-Jones & Brown, 1981; Hettema, Prescott, Myers, Neale, & Kendler, 2005; Waszczuk, Zavos, Gregory, & Eley, 2014), and some suggest that incorporating disorder-specific approaches to psychological treatment can improve outcomes (Clark & Beck, 2011). Studies focused on expanding sample sizes may find that meta-analyses combining data from cohorts ascertained with self-reported or symptom-based diagnoses are sufficient to identify effects that are shared between anxiety and major depressive disorders (Hettema et al., 2005; Morneau-Vaillancourt et al., 2020). However, in order to understand disorder-specific risk factors or investigate treatment approaches for these disorders, particularly the anxiety subtypes, findings may vary depending on the ascertainment method used in the study. The population of interest should therefore be considered when selecting measures for future studies. Both methods may have an important role in future research depending on the aims of individual studies. Those focused on increasing participation and reducing the time burden for participants and researchers may consider the use of self-reported measures, taking into account the differences by disorder in sensitivity and specificity for symptom-based diagnoses. For instance, a study recruiting from clinical populations may use self-reported MDD to ascertain MDD diagnosis. In contrast, those particularly interested in identifying cases with specific anxiety disorder subtypes are likely to benefit from use of the symptom-based approach.

These results can further be considered in the context of the efficacy of self-reported, broad diagnostic measures, which have been explored the most thus far with regard to depression. There is a diversity of opinions concerning the value and utility of these brief measures in the field, especially in large-scale research such as psychiatric genetics. For example, researchers have found that "broad" depression (e.g., depression defined using self-reported diagnoses or self-reports of treatment seeking) is non-specific to MDD and has lower heritability estimates than symptom-based MDD (Cai et al., 2020; Glanville et al., 2020). Misclassification dilutes the power of case-control analyses to detect differences between the samples (Manchia et al., 2013; Schork et al., 2018). As such, the lower heritability estimates of self-reported measures of MDD may indicate that this approach has a higher rate of

misclassification for true cases and controls than the symptom-based diagnoses. Indeed, self-reported MDD, any anxiety, and GAD in this study had high agreement with symptom-based diagnoses when identifying cases, but differed in the classification of those without a diagnosis. It has been argued that symptom-based measures are preferable over their self-reported counterparts, and that studies utilising self-reported diagnoses for the purpose of case-control comparisons are more likely to identify effects that are non-specific, complicating efforts to disentangle disorder-specific factors and treatments (Cai et al., 2020; Phillips & Kendler, 2021). For studies that are unable to administer symptom-based assessments or existing studies that did not include these measures, combining multiple broad diagnostic measures (e.g., self-reported diagnoses, self-reported 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 symptom-based MDD (Glanville et al., 2020).

In terms of the differences observed in the categorisation of the anxiety disorders, the lower proportion of self-reported 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 (McManus et al., 2016; Rayner et al., 2019) and those that do more commonly discuss their problems with a GP rather than a mental health professional (McManus et al., 2016). Research has shown that there is an under-recognition of anxiety disorders, particularly by GPs (Arikian & Gorman, 2001; Fernández et al., 2012; Tylee & Walters, 2007; Vermani, Marcus, & Katzman, 2011). GPs have limited amounts of time and resources and lack specialised training to conduct comprehensive assessments of anxiety symptoms (Baird, Charles, Honeyman, Maguire, & Das, 2016). It is therefore possible that GPs encountering distressed patients may identify symptoms as “anxiety” without specifying a disorder. Notably, in the GLAD and COPING NBR studies, the phrasing of the self-reported GAD item encapsulates general nerves or anxiety to account for this, which may have resulted in an overestimate of the number of participants given a GAD diagnosis.

Our finding that self-reported panic attacks had higher agreement than self-reported panic disorder to symptom-based panic disorder could be further indication of this under-recognition. Although panic attacks are not specific to panic disorder, they are more recognisable and straightforward to diagnose than panic disorder. However, studies have found that the majority of individuals who experience a panic attack do not have panic disorder (Kessler et al., 2006). Consequently, this finding could instead indicate a lack of specificity of the symptom-based panic disorder measure.

4.4. Conclusion

Large-scale research projects that lack the resources to conduct “gold standard” clinical interviewing commonly utilise questionnaires applying symptom-based or self-reported diagnostic methods. We compared these two approaches and found good comparability between symptom-based and self-reported MDD and "any anxiety" disorder for categorisation of participants with a diagnosis, although performance varied between the case and general population cohorts in this study. Ascertainment of participants with diagnoses for the individual anxiety disorders was largely different depending on which phenotyping measure was applied. Taking our results together with previous studies, we suggest that self-reported diagnoses may be sufficient depending on the aims of the research and the population under study, but may not be suitable for case-control studies investigating disorder-specific risk factors or outcomes. Notably, prior research provides little insight regarding the validity of self-reported or symptom-based diagnoses against clinical interviews. The differences observed in this study highlighted the need for further validation of these diagnoses against clinical interviews to advise measure selection and ensure translatability of research incorporating these measures.

Funding

This work was supported by the National Institute for Health Research (NIHR) BioResource [RG94028, RG85445], National Institute for Health Research (NIHR) Biomedical Research Centre [IS-BRC-1215-20018], Health and Social Care Research and Development (HSC R&D) Division, Public Health Agency [COM/5516/18], Medical Research Council (MRC) Mental Health Data Pathfinder Award (MC_PC_17,217), and the National Centre for Mental Health funding through Health and Care Research Wales. Profs Eley and Breen are part-funded by a program grant from the UK Medical Research Council (MR/V012878/1). Dr Buckman was supported by a Clinical Research Fellowship from the Wellcome Trust (201292/Z/16/Z). Dr. Goldsmith receives funding from the National Institute for Health Research (NIHR), Medical Research Council (MRC), National Institutes of Health (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). Dr Hübel acknowledges funding from Lundbeckfonden (R276-2018-4581).

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.

Acknowledgement

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, 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.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.janxdis.2021.102491.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Amer Psychiatric Pub.

Arikian, S. R., & Gorman, J. M. (2001). A review of the diagnosis, pharmacologic treatment, and economic aspects of anxiety disorders. *Primary Care Companion to The Journal of Clinical Psychiatry, 3*(3), 110–117 (Jun).

Baird, B., Charles, A., Honeyman, M., Maguire, D., & Das, P. (2016). *Understanding pressures in general practice*. In K. O’Neill (Ed.). The King’s Fund (May).

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience, 17*(3), 327–335 (Sep).

Brugha, T. S., Bebbington, P. E., & Jenkins, R. (1999). A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychological Medicine, 29*(5), 1013–1020 (Sep).
Byrne, E. M., Kirk, K. M., Medland, S. E., McGrath, J. J., Colodro-Conde, L., Parker, R., … Martin, N. G. (2020). Cohort profile: The Australian genetics of depression study. *BMJ Open, 10*(5), Article e032580. May 26.
Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics, 52*(4), 437–447 (Apr).
Carlbring, P., Forslin, P., Ljungstrand, P., Willebrand, M., Strandlund, C., Ekselius, L., & Andersson, G. (2002). Is the Internet-administered CIDI-SF Equivalent to a Clinician-administered SCID Interview? *Cognitive Behavioral Therapy, 31*(4), 183–189 (Jan).
Clark, D. A., & Beck, A. T. (2011). *Cognitive Therapy of Anxiety Disorders: Science and Practice*. New York: The Guilford Press (Updated).
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46 (Apr 1).
Davies, M. R., Kalsi, G., Armour, C., Jones, I. R., McIntosh, A. M., Smith, D. J., … Breen, G. (2019). The Genetic Links to Anxiety and Depression (GLAD) Study: Online recruitment into the largest recontactable study of depression and anxiety. *Behaviour Research and Therapy, 123*, Article 103503 (Oct 24).
Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., … Hotopf, M. (2020). Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. *BJPsych Open, 6*(2), Article e18 (Feb 6).
Davis, K. A. S., Cullen, B., Adams, M., Brailean, A., Breen, G., Coleman, J. R. I., … Hotopf, M. (2019). Indicators of mental disorders in UK Biobank-A comparison of approaches. *International Journal of Methods in Psychiatric Research, 28*(3), Article e1796 (Aug 8).
Fernández, A., Rubio-Valera, M., Bellón, J. A., Pinto-Meza, A., Luciano, J. V., Mendive, J. M., … DASMAP, I. (2012). Recognition of anxiety disorders by the general practitioner: results from the DASMAP study. *General Hospital Psychiatry, 34*(3), 227–233 (Jun).
Finlay-Jones, R., & Brown, G. W. (1981). Types of stressful life event and the onset of anxiety and depressive disorders. *Psychological Medicine, 11*(4), 803–815 (Nov).
First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5—Research version (SCID-5)*. Arlington, VA: American Psychiatric Association.
Fleet, R. P., Dupuis, G., Marchand, A., Burelle, D., Arsenault, A., & Beitman, B. D. (1996). Panic disorder in emergency department chest pain patients: prevalence, comorbidity, suicidal ideation, and physician recognition. *The American Journal of Medicine, 101*(4), 371–380 (Oct).
Glanville, K. P., Coleman, J. R., Howard, D. M., Pain, O., Hanscombe, K. B., Jermy, B., & Lewis, C. M. (2020). Multiple measures of depression to enhance validity of Major Depressive Disorder in the UK Biobank. *medRxiv* (Sep 22).
Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American Journal of Psychiatry, 158*(10), 1568–1578 (Oct).
Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *The Archives of General Psychiatry, 62*(2), 182–189 (Feb).
Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience, 22*(3), 343–352 (Mar).
Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., … Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics, 48*(9), 1031–1036 (Sep).
Katerndahl, D. A., & Realini, J. P. (1995). Where do panic attack sufferers seek care? *Journal of Family Practice* (Mar 1).
Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *The American Journal of Psychiatry, 163*(1), 109–114 (Jan).
Kessler, D., Bennewith, O., Lewis, G., & Sharp, D. (2002). Detection of depression and anxiety in primary care: follow up study. *BMJ, 325*(7371), 1016–1017 (Nov 2).
Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H.-U. (1998). The World Health Organization composite international diagnostic interview short-form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7*(4), 171–185 (Nov).
Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General, 62*(6), 593–602 (Jun).
Kessler, R. C., Chiu, W. T., Jin, R., Ruscio, A. M., Shear, K., & Walters, E. E. (2006). The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. *The Archives of General Psychiatry, 63*(4), 415–424 (Apr).
Kuhn M. caret: Classification and Regression Training. [Internet]. 2020 [cited 2021 May 6]. Available from: ⟨https://CRAN.R-project.org/package=caret⟩.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174 (Mar).
Levinson, D., Potash, J., Mostafavi, S., Battle, A., Zhu, X., & Weissman, M. (2017). Brief assessment of major depression for genetic studies: Validation of Cidi-Sf screening with scid interviews. *European Neuropsychopharmacology, 27*, S448.
Lynch, P., & Galbraith, K. M. (2003). Panic in the emergency room. *The Canadian Journal of Psychiatry, 48*(6), 361–366 (Jul).
Manchia, M., Cullis, J., Turecki, G., Rouleau, G. A., Uher, R., & Alda, M. (2013). The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS ONE, 8*(10), Article e76295 (Oct 11).
McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). *Health and wellbeing in England: Adult psychiatric morbidity survey 2014*. Leeds: NHS Digital (Jan).
McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153–157 (Jun).
Morneau-Vaillancourt, G., Coleman, J. R. I., Purves, K. L., Cheesman, R., Rayner, C., Breen, G., & Eley, T. C. (2020). The genetic and environmental hierarchical structure of anxiety and depression in the UK Biobank. *Depression and Anxiety, 37*(6), 512–520 (Jan 17).
Patten, S. B. (1997). Performance of the Composite International Diagnostic Interview Short Form for major depression in community and clinical samples. *Chronic Diseases in Canada, 18*(3), 109–112.
Phillips, M. L., & Kendler, K. S. (2021). Three important considerations for studies examining pathophysiological pathways in psychiatric illness: In-depth phenotyping, biological assessment, and causal inferences. *JAMA Psychiatry, 78*, 697–698 (Mar 17).
Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., … Eley, T. C. (2020). A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry, 25*(12), 3292–3303 (Dec).
R Core Team. (2018). *R: A language and environment for statistical computing [Internet]*. Vienna, Austria: R Foundation for Statistical Computing ([cited 2019 Sep 20]. Available from:) https://www.r-project.org/.
Rayner, C., Coleman, J. R. I., Purves, K. L., Cheesman, R., Hübel, C., Gaspar, H., … Eley, T. C. (2019). Genetic influences on treatment-seeking for common mental health problems in the UK biobank. *Behaviour Research and Therapy, 121*, Article 103413 (Jun 15).
Sanchez-Villegas, A., Schlatter, J., Ortuno, F., Lahortiga, F., Pla, J., Benito, S., & Martinez-Gonzalez, M. A. (2008). 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, 8*, 43. Jun 17.
Schork, A., Hougaard, D., Nordentoft, M., Mors, O., Boerglum, A., Bo Mortensen, P., & Werge, T. (2018). 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* (Dec 4).
Shimada-Sugimoto, M., Otowa, T., & Hettema, J. M. (2015). Genetics of anxiety disorders: Genetic epidemiological and molecular studies in humans. *Psychiatry and Clinical Neurosciences, 69*(7), 388–401 (Jul).
Stuart, A. L., Pasco, J. A., Jacka, F. N., Brennan, S. L., Berk, M., & Williams, L. J. (2014). Comparison of self-report and structured clinical interview in the identification of depression. *Comprehensive Psychiatry, 55*(4), 866–869 (May).
Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *The American Journal of Psychiatry, 157*(10), 1552–1562 (Oct).
Tylee, A., & Walters, P. (2007). Underrecognition of anxiety and mood disorders in primary care: Why does the problem exist and what can be done? *The Journal of Clinical Psychiatry, 68*(Suppl 2), 27–30.
van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nature Reviews Genetics, 20*(10), 567–581 (Oct).
Vermani, M., Marcus, M., & Katzman, M. A. (2011). Rates of detection of mood and anxiety disorders in primary care: A descriptive, cross-sectional study. *Primary Care Companion for CNS Disorders, 13*(2).
Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. *The Archives of General Psychiatry, 62*(6), 629–640 (Jun).
Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M., & Eley, T. C. (2014). The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. *JAMA Psychiatry, 71*(8), 905–916 (Aug).
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *JOSS, 4*(43), 1686 (Nov 21).
World Health Organization. (1990). *Composite international diagnostic interview*. Geneva, Switzerland: World Health Organization.
World Health Organization. (2017). *Depression and Other Common Mental Disorders: Global Health Estimate [Internet]*. Geneva: World Health Organization ([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⟩.
Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Viktorin, A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*(5), 668–681 (May).