Original Article

![CAMBRIDGE UNIVERSITY PRESS]

How much or how often? Examining the screening properties of the DSM cross-cutting symptom measure in a youth population-based sample

João Pedro Gonçalves Pacheco[1,2,3] (iD), Christian Kieling[2,4,5] (iD), Pedro H. Manfro[2] (iD), Ana M. B. Menezes[6], Helen Gonçalves[6] (iD), Isabel O. Oliveira[6], Fernando C. Wehrmeister[6,7] (iD), Luis Augusto Rohde[5,8,9] (iD) and Maurício Scopel Hoffmann[1,2,3,10] (iD)

[1]Department of Neuropsychiatry, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil; [2]Graduate Program in Psychiatry and Behavioral Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil; [3]Mental Health Epidemiology Group, Santa Maria, Rio Grande do Sul, Brazil; [4]Department of Psychiatry and Legal Medicine, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil; [5]ADHD Outpatient Program & Developmental Psychiatry Program, Hospital de Clínicas de Porto Alegre, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil; [6]Postgraduate Program in Epidemiology, Universidade Federal de Pelotas, Pelotas, RS, Brazil; [7]Rady Faculty of Health Sciences, Institute for Global Public Health, University of Manitoba, Winnipeg, Manitoba, Canada; [8]National Institute of Developmental Psychiatry for Children and Adolescents (INCT-CNPq), São Paulo, Brazil; [9]UniEduK, Indaiatuba, São Paulo, Brazil and [10]Care Policy and Evaluation Centre, London School of Economics and Political Science, London, UK

Abstract

Background. The DSM Level 1 Cross-Cutting Symptom Measure (DSM-XC) allows for assessing multiple psychopathological domains. However, its capability to screen for mental disorders in a population-based sample and the impact of adverbial framings (intensity and frequency) on its performance are unknown.

Methods. The study was based on cross-sectional data from the 1993 Pelotas birth cohort in Brazil. Participants with completed DSM-XC and structured diagnostic interviews ($n = 3578$, aged 22, 53.6% females) were included. Sensitivity, specificity, positive (LR+), and negative (LR−) likelihood ratios for each of the 13 DSM-XC domains were estimated for detecting five internalizing disorders (bipolar, generalized anxiety, major depressive, post-traumatic stress, and social anxiety disorders) and three externalizing disorders (antisocial personality, attention-deficit/hyperactivity, and alcohol use disorders). Sensitivities and specificities >0.75, LR+ > 2 and LR− < 0.5 were considered meaningful. Values were calculated for the DSM-XC's original scoring and for adverbial framings.

Results. Several DSM-XC domains demonstrated meaningful screening properties. The anxiety domain exhibited acceptable sensitivity and LR− values for all internalizing disorders. The suicidal ideation, psychosis, memory, repetitive thoughts and behaviors, and dissociation domains displayed acceptable specificity for all disorders. Domains also yielded small but meaningful LR+ values for internalizing disorders. However, LR+ and LR− values were not generally meaningful for externalizing disorders. Frequency-framed questions improved screening properties.

Conclusions. The DSM-XC domains showed transdiagnostic screening properties, providing small but meaningful changes in the likelihood of internalizing disorders in the community, which can be improved by asking frequency of symptoms compared to intensity. The DSM-XC is currently lacking meaningful domains for externalizing disorders.

Introduction

There is a growing clinical and academic interest in investigating symptoms that cut across multiple mental disorders (Eaton et al., 2023; Forbes et al., 2024; Gibbons, Farmer, Shaw, & Chung, 2023), such as anxiety, insomnia, depressed mood, hallucinations, among others. Both authoritative texts (American Psychiatric Association, 2013) and empirical studies (Forbes, Tackett, Markon, & Krueger, 2016; Waszczuk et al., 2017) have encouraged dimensional approaches to psychopathology assessment, with the aim of elucidating shared features among different disorders. An emerging tool is The Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) Level 1 Cross-Cutting Symptom Measure (DSM-XC), which consists of 23 self-reported questions that explore 13 different mental health domains. However, the current body of evidence supporting the utility of cross-cutting symptom

assessments, including the DSM-XC, remains limited. This gap presents an unexplored opportunity not only for understanding the connections between different diagnostic categories but also for optimizing community-based mental disorder screening.

The U.S. Preventive Services Task Force (USPSTF) supports the use of disorder-specific scales in screening recommendations for anxiety (US Preventive Services Task Force et al., 2023a), depression (US Preventive Services Task Force et al., 2023b), and unhealthy drug and alcohol use among adults (US Preventive Services Task Force et al., 2018, 2020). Notable scales presenting good diagnostic accuracy for these purposes (Mulvaney-Day et al., 2018) include the Patient Health Questionnaire-9 (PHQ-9, sensitivity = 0.85, specificity = 0.85) and Center for Epidemiological Studies Depression (CES-D, sensitivity = 0.83, specificity = 0.78) for major depression (Negeri et al., 2021; Vilagut, Forero, Barbaglia, & Alonso, 2016), the General Anxiety Disorder-7 (GAD-7) for generalized anxiety disorder (sensitivity = 0.83, specificity = 0.84) (Plummer, Manea, Trepel, & McMillan, 2016), the National Institute on Alcohol Abuse and Alcoholism (NIAAA) for current alcohol use disorder (sensitivity = 0.88, specificity = 0.67) (Smith, Schmidt, Allensworth-Davies, & Saitz, 2009), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) for tobacco, alcohol, and cannabis use disorders (sensitivities = 0.95–1.00; specificities = 0.79–0.93) (Gryczynski et al., 2015), and the Tobacco, Alcohol, Prescription Medication and Other Substance Use (TAPS, sensitivities = 0.70–0.74, specificities 0.85–0.95 for tobacco, alcohol and marijuana) (McNeely et al., 2016), among others. Nevertheless, current recommendations lack the inclusion of multi-disorder tools. Only a handful of such screening instruments are available, such as the PHQ-4, The Hospital Anxiety and Depression Scale (HADS), and the Mental Health Index (MHI-5) (Means-Christensen et al., 2005). The PHQ-4 (Löwe et al., 2010) is an ultra-short assessment that combines the PHQ-2 (sensitivity = 0.89, specificity = 0.75) (Mitchell, Yadegarfar, Gill, & Stubbs, 2016) and GAD-2 (sensitivity = 0.80, specificity = 0.81) (Plummer et al., 2016), facilitating simultaneous screening for depression and anxiety. The HADS is a lengthier but commonly used tool for identifying depression and anxiety in medically ill patients, with sensitivities and specificities for depression and anxiety subscales of approximately 0.80 (Bjelland, Dahl, Haug, & Neckelmann, 2002). However, the extent of diagnostic insight provided by comprehensive assessments remains largely unexplored for psychiatric disorders.

The symptoms included in the DSM-XC exhibit similarities to the PHQ-2 (Löwe, Kroenke, & Gräfe, 2005; Mitchell et al., 2016) and the GAD-2 (Plummer et al., 2016). However, the utility of these scales has been ascertained only for the specific diagnoses they were designed to screen for. Given the recognized high comorbidity among mental disorders (Plana-Ripoll et al., 2019) and the occurrence of overlapping symptoms across diagnoses (Forbes et al., 2024), it is pertinent to consider expanding the use of these brief screeners to identify multiple disorders. In addition, a comprehensive examination of the screening properties of the 13 DSM-XC domains can help pinpoint which cross-cutting symptoms are more effective in either confirming or ruling out specific diagnoses.

Previous studies suggested the DSM-XC (Bastiaens & Galus, 2018; Doss & Lowmaster, 2022; Mahoney et al., 2020) could be used for screening, but they lacked the use of diagnostic interviews and assessed convenience samples with limited generalizability and ecological validity. Additionally, the DSM-XC merges questioning about the intensity ('how much') and frequency ('how often') of symptoms, which could impact its psychometric and screening properties (Gong et al., 2013; Krabbe & Forkmann, 2014; Krishnakumar et al., 2021; Levin, Schneider, & Gaeth, 1998), but empirical studies on the impact on screening utility of different framings are not known.

Our aim was to assess the screening properties (sensitivity, specificity, and likelihood ratios [LRs]) of the DSM-XC's domains to detect eight mental disorders, representative of both internalizing and externalizing spectra, in the community. We will address this by assessing the instruments' original operationalization merging intensity and frequency of symptoms and conducting an analysis for both framings separately.

Methods

Assessments

Participants

We analyzed data from the latest assessment of 22 years of an ongoing longitudinal population-based study of the city of Pelotas, Brazil, that recruited all children born alive in 1993 (Victora et al., 2008), with a retention rate of 76.3%. Detailed information on the cohort is described elsewhere (Gonçalves et al., 2018; Victora et al., 2006). Only individuals with completed DSM-XC and MINI data were included in the present study. Written informed consent was obtained from all participants. The study was approved by the Institutional Review Board of the School of Medicine, Universidade Federal de Pelotas (approval number 1.250.366).

The DSM-5 Level 1 cross-cutting symptom measure

The DSM-XC is a self-reported measure that examines different mental health domains to assist clinicians in identifying comorbidity that could impact a patient's treatment and prognosis (American Psychiatric Association, 2013). The cross-cutting measures for adults involve two levels of inquiry. At Level 1, the questionnaire consists of 23 items distributed in 13 domains: depression, anger, mania, anxiety, somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviors, dissociation, personality functioning, and substance use. Items ask, 'how much (or how often) have you been bothered by' a given problem during the previous two weeks and provide a 5-point Likert scale ranging from 0 = 'None (not at all)' to 4 = 'Severe (nearly every day)'. A rating of mild (i.e. 2) or greater on any item within a domain, except for substance use, suicidal ideation, and psychosis is considered a positive screen. For substance use, suicidal ideation, and psychosis, a rating of slight (i.e. 1) or greater on any item within the domain is considered a positive domain. The threshold scores in any of its domains indicate the need for a thorough clinical investigation with the respective Level 2 Cross-Cutting Symptom Measure. Level 2 questions offer a more detailed assessment of specific domains. Except for suicidal ideation, psychosis, memory problems, dissociation, and personality functioning, each domain in the adult version of the DSM-XC corresponds to a Level 2 measure, which contains more questions about each screened disorder. Additional details on Level 1 and Level 2 measures can be found online at www.psychiatry.org/dsm5.

In the original DSM-XC, there are two simultaneous framings when asking about symptoms: intensity (i.e. how much) and frequency (i.e. how often), requiring the respondent to choose a

single answer. After a pilot test in Brazil, we identified that some participants had difficulty understanding the double-barreled questions, which led to the decision to split each question into two, and answers were recorded on two separate 5-point scales: one for intensity and another for frequency. To compute the score as intended by the original DSM-XC 'OR' rule, the highest score in each question was considered when analyzed jointly. For example, an answer of 0 (none) for how much and 2 (several days) for how often, or vice versa, was coded as 2. An additional adaptation was made to the instrument by dividing the last item (item 23) into two (items 23 and 24, as shown in online Supplementary Table S1) to allow for the exploration of substance use and self-prescription/misuse separately.

Assessment of psychiatric diagnoses

Participants were also assessed for the presence of psychiatric disorders with an adapted Brazilian version of the MINI International Neuropsychiatric Interview (MINI) (Amorim, 2000), administered by trained clinical psychologists, who did not have access to the DSM-XC responses. The MINI 5.0.0 is a structured and standardized diagnostic interview based on the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 1994), and was used to assess the following disorders: antisocial personality disorder (APD), attention-deficit/hyperactivity disorder (ADHD), bipolar disorder (BD), generalized anxiety disorder (GAD), major depressive disorder (MDD), post-traumatic stress disorder (PTSD) and social anxiety disorder (SAD). Moreover, to capture the DSM's impairment criterion, we included an additional question inquiring participants how much impairment the reported symptoms caused in their life, with response options of none, mild, moderate, or severe (Manfro et al., 2021; Matte et al., 2015). The presence of clinical impairment was operationally defined as having a score of moderate or severe. The Alcohol Use Disorders Identification Test (AUDIT), a 10-question instrument, was used to identify participants with active alcohol use disorder (AUD). A score of ≥ 15 for men and ≥ 13 for women was considered positive. The AUDIT validity has been previously investigated in Brazil (Meneses-Gaya et al., 2010).

Data analyses

We followed the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 guidelines (Bossuyt et al., 2015), reported in online Supplementary Table S2.

Screening properties

We estimated the number of true positives, false positives, true negatives, and false negatives for each of the 13 DSM-XC domains to detect individuals with psychiatric disorders. We then estimated the sensitivity, specificity, positive (LR+), and negative (LR−) likelihood ratios (Altman & Bland, 1994; Grimes & Schulz, 2005) for each domain-diagnosis pair.

Sensitivities and specificities higher than 0.75 (Power, Fell, & Wright, 2013), LR+ values higher than 2, and LR− values lower than 0.5 were considered clinically useful (Grimes & Schulz, 2005; Jaeschke, Guyatt, & Sackett, 1994). Exact binomial 95% confidence intervals (95% CI) were calculated for sensitivity and specificity (Collett, 2002). Confidence intervals for LR+ and LR− values are based on formulae provided by Simel, Samsa, and Matchar (1991).

Analysis of variability in screening properties by framing

We first described the by-item response differences and calculated the polychoric correlation and Cohen's kappa to measure the reliability between frames for each DSM-XC item. Second, we calculated and compared all screening properties between framings. Overlapping confidence intervals were considered statistically equivalent. Paired Wilcoxon tests with DSM-XC domains as dependent variables were conducted to compare the differences in screening properties between the two framings. Screening properties of the two framings were also compared with the OR rule (reference group) with paired Wilcoxon tests with DSM-XC domains as dependent variables.

Logistic regression models were used to test potential interactions between intensity and frequency that could indicate an advantage of a particular framing. The odds ratios for specific combinations, namely Intensity (−)/Frequency (+), Intensity (+)/Frequency (−), and Intensity (+)/Frequency (+) were calculated with Intensity (−)/Frequency (−) as the reference group for each domain with a corresponding disorder.

We also tested if clinical impairment could be a factor for potential differences in framings instead of the framing effect itself. Because of this, secondary analyses were conducted by removing the clinical impairment criteria. The diagnosis of AUD was based on AUDIT scores and was not included in this analysis.

All analyses were performed using the software R, version 4.1.3. Screening properties were evaluated using the epiR package, version 2.0.44 (Stevenson et al., 2020). Paired Wilcoxon tests were conducted using the rstatix package, version 0.7.0 (Kassambara, 2021). Code is available at http://osf.io/q54yu.

Results

Sample description

A total of 3578 subjects with completed DSM-XC and MINI data participated in the 2015 follow-up study. The mean age was 22.57 years. Participants were mainly of the female sex (53.58%). Most individuals had working experience (94.10%), 63.30% had finished high school, and 84.13% were single. At least one psychiatric diagnosis was present in 813 subjects (22.72%). Sample characteristics are provided in Table 1. The prevalence of specific diagnoses ranged from 1.00% (APD) to 10.56% (GAD) (Table 1). A positive screen in the anxiety domain was observed in 50% of participants, followed by substance use at 47% and depression at 43%. Other prevalent domains included somatic symptoms (36%), anger (32%), and personality functioning (30%). Suicidal ideation (9%), psychosis (14%), and dissociation (16%) were the least frequent domains. Full details of the number and proportion of positive screens for each domain is provided in online Supplementary Table S3. Cross-tabulation of the DSM-XC domains by the results of the reference standards (MINI diagnoses and AUDIT) are provided in online Supplementary Tables S4–S19.

Sensitivity

The anxiety domain demonstrated excellent sensitivity across all internalizing disorders, with sensitivity values ranging from 0.86 to 0.95 (Table 2). In particular, the depression domain exhibited good sensitivity for both GAD (0.80, 95% CI 0.76–0.84) and MDD (0.95, 95% CI 0.89–0.98). The personality functioning domain showed adequate sensitivity for MDD (0.86, 95% CI 0.78–0.92). For externalizing disorders, only the substance use domain displayed good sensitivity for AUD (0.87, 95% CI 0.81–0.91).

Table 1. Sample description

Sample characteristics	Sample (*n* = 3578)
Age, mean (s.d.), y	22.57 (0.33)
Sex	
Female, *n* (%)	1917 (53.58)
NEET, *n* (%)	716 (20.85)
Missing, *n* (%)	66 (1.84)
Attainment	
Incomplete primary or illiterate, *n* (%)	61 (1.70)
Primary, *n* (%)	1248 (34.88)
Secondary, *n* (%)	2065 (57.71)
Tertiary, *n* (%)	200 (5.59)
Missing, *n* (%)	4 (0.11)
Family income per capita	
BRL 0 to 1254, *n* (%)	2584 (72.22)
BRL 1255 to 2004, *n* (%)	522 (14.59)
BRL 2005 to 8640, *n* (%)	287 (8.02)
BRL > 8641, *n* (%)	6 (0.17)
Missing, *n* (%)	179 (5.00)
Ever worked, *n* (%)	3367 (94.10)
Currently working, *n* (%)	2267 (63.36)
Missing, *n* (%)	211 (5.90)
Marital status	
Married	548 (15.32)
Divorced	19 (0.53)
Single	3010 (84.13)
Widowed	1 (0.00)
Lifetime use of tobacco, *n* (%)	975 (27.25)
Lifetime use of alcoholic beverages, *n* (%)	3397 (94.94)
Prevalence of current psychiatric diagnoses	
Alcohol use disorder, *n* (%)	208 (5.81)
Antisocial personality disorder, *n* (%)	36 (1.00)
Attention-deficit/hyperactivity disorder, *n* (%)	162 (4.53)
Bipolar disorder, *n* (%)	60 (1.68)
Generalized anxiety disorder, *n* (%)	378 (10.56)
Major depressive disorder, *n* (%)	105 (2.93)
Post-traumatic stress disorder, *n* (%)	155 (4.33)
Social anxiety disorder, *n* (%)	178 (4.97)
Any diagnosis, *n* (%)	813 (22.72)

Note: NEET, Not in Education, Employment, or Training; BRL, Brazilian *Real* (Currency Unit).

Specificity

Various domains, including suicidal ideation, psychosis, memory, repetitive thoughts and behaviors, and dissociation, exhibited good specificity across all tested disorders (Table 2). Specifically, suicidal ideation had a specificity range of 0.91–0.92, psychosis ranged from 0.86 to 0.87, memory from 0.83 to 0.86, repetitive thoughts and behaviors from 0.81 to 0.84, and dissociation from 0.84 to 0.87. Sleep problems showed good specificity for both GAD (0.79, 95% CI 0.77–0.80) and PTSD (0.77, 95% CI 0.76–0.78).

Likelihood ratios

Differences in LR+ values were observed between internalizing and externalizing disorders (Table 3). Several domains, namely suicidal ideation, repetitive thoughts and behaviors, dissociation, and personality functioning, exhibited clinically meaningful LR+ values for all internalizing disorders, but not for externalizing disorders.

The anxiety domain demonstrated meaningful LR− values for all internalizing disorders, ranging from 0.09 to 0.27. The depression domain also displayed good LR− values for most internalizing disorders, specifically GAD, MDD, and PTSD. Other domains, such as anger, personality functioning, and repetitive thoughts and behaviors showed meaningful LR− values for specific disorders. For externalizing disorders, only the substance use domain presented acceptable LR− values for AUD.

Variability in above-threshold domains and screening properties by adverbial framing

Polychoric correlation coefficients between 'intensity' and 'frequency' framings ranged from 0.77 to 0.90 and Kappa coefficient ranged from 0.50 to 0.66. Detailed results are shown in online Supplementary Table S1, and the item response distribution for the DSM-XC is shown in online Supplementary Fig. S1.

Intensity-framed questions resulted in a similar number of above-threshold domains compared to the original rule (online Supplementary Table S2). Frequency-framed questions showed a lower number of above-threshold domains, which was accompanied by significant improvements in LR+ values (Fig. 1) mainly for the depression, anger, and anxiety domains. LR− values were not significantly impacted by the framing variation (Fig. 2). Wilcoxon tests comparing framings across all DSM-XC domains showed that intensity-framed questions present marginally better sensitivity and LR− values, but worse specificity, and LR+ values when compared to frequency-framed questions (online Supplementary Figs S2 and S3, Supplementary Tables S20–S23). The intensity framing showed minor differences in sensitivity, specificity, and LR+ when compared to the OR rule, while the frequency framing showed lower sensitivity, but higher specificity and LR+ compared to the OR rule. Differences in LR− were very small for both framings compared to the OR rule (online Supplementary Tables S24–S27).

Logistic regressions of specific framing combinations (online Supplementary Table S28) indicate that for mania-BD, personality functioning-APD, and substance use-AUD pairs, the Intensity (−)/Frequency (+) group had a significant higher odds compared to the reference group, while the Intensity (+)/Frequency (−) had not. For the depression domain to detect MDD and anxiety domain to detect GAD and SAD, all operationalizations had a higher likelihood of the corresponding disorder.

Full details of the screening properties by framing are described in online Supplementary Tables S29–S52.

Secondary analyses showed that the findings remained stable even without the MINI clinical impairment criterion (online Supplementary Figs S4–S7, Supplementary Tables S53–S78).

Table 2. Sensitivity and specificity of DSM-XC domains to detect eight psychiatric disorders

Accuracy metric, 95% CI	Domain	Internalizing disorders					Externalizing disorders		
		Generalized anxiety disorder	Social anxiety disorder	Major depressive disorder	Post-traumatic stress disorder	Bipolar disorder	Attention-deficit/ hyperactivity disorder	Alcohol use disorder	Antisocial personality disorder
Sensitivity	Depression	**0.80 (0.76–0.84)**	0.77 (0.70–0.83)	**0.95 (0.89–0.98)**	0.79 (0.72–0.85)	0.77 (0.64–0.87)	0.69 (0.61–0.76)	0.59 (0.52–0.66)	0.58 (0.41–0.74)
	Anger	0.70 (0.65–0.74)	0.61 (0.53–0.68)	0.76 (0.67–0.84)	0.72 (0.64–0.79)	0.75 (0.62–0.85)	0.59 (0.51–0.67)	0.50 (0.43–0.57)	0.44 (0.28–0.62)
	Mania	0.43 (0.38–0.48)	0.46 (0.38–0.53)	0.46 (0.36–0.56)	0.55 (0.47–0.63)	0.65 (0.52–0.77)	0.43 (0.35–0.51)	0.40 (0.34–0.47)	0.42 (0.26–0.59)
	Anxiety	**0.87 (0.83–0.90)**	**0.86 (0.80–0.91)**	**0.95 (0.89–0.98)**	**0.94 (0.88–0.97)**	**0.95 (0.86–0.99)**	0.75 (0.67–0.81)	0.63 (0.57–0.70)	0.69 (0.52–0.84)
	Somatic Symptoms	0.67 (0.62–0.71)	0.60 (0.53–0.67)	0.75 (0.66–0.83)	0.72 (0.65–0.79)	0.78 (0.66–0.88)	0.54 (0.46–0.62)	0.49 (0.42–0.56)	0.50 (0.33–0.67)
	Suicidal ideation	0.20 (0.16–0.25)	0.25 (0.19–0.32)	0.38 (0.29–0.48)	0.27 (0.20–0.35)	0.33 (0.22–0.47)	0.19 (0.13–0.25)	0.15 (0.10–0.20)	0.11 (0.03–0.26)
	Psychosis	0.28 (0.23–0.32)	0.30 (0.24–0.38)	0.34 (0.25–0.44)	0.39 (0.32–0.48)	0.53 (0.40–0.66)	0.29 (0.22–0.37)	0.29 (0.23–0.36)	0.39 (0.23–0.57)
	Sleep problems	0.54 (0.49–0.59)	0.56 (0.48–0.63)	0.69 (0.59–0.77)	0.65 (0.57–0.73)	0.57 (0.43–0.69)	0.47 (0.39–0.55)	0.48 (0.41–0.55)	0.39 (0.23–0.57)
	Memory	0.38 (0.33–0.43)	0.42 (0.34–0.49)	0.39 (0.30–0.49)	0.46 (0.38–0.55)	0.48 (0.35–0.62)	0.37 (0.30–0.45)	0.28 (0.22–0.35)	0.25 (0.12–0.42)
	Repetitive thoughts and behaviors	0.49 (0.44–0.54)	0.46 (0.38–0.53)	0.63 (0.53–0.72)	0.68 (0.60–0.76)	0.68 (0.55–0.80)	0.42 (0.34–0.50)	0.38 (0.31–0.44)	0.42 (0.26–0.59)
	Dissociation	0.40 (0.35–0.45)	0.39 (0.32–0.47)	0.58 (0.48–0.68)	0.49 (0.41–0.57)	0.60 (0.47–0.72)	0.36 (0.29–0.44)	0.30 (0.24–0.37)	0.42 (0.26–0.59)
	Personality functioning	0.63 (0.58–0.68)	0.62 (0.54–0.69)	**0.86 (0.78–0.92)**	0.75 (0.68–0.82)	0.72 (0.59–0.83)	0.52 (0.44–0.60)	0.53 (0.46–0.60)	0.61 (0.43–0.77)
	Substance use	0.59 (0.54–0.64)	0.60 (0.52–0.67)	0.63 (0.53–0.72)	0.65 (0.56–0.72)	0.67 (0.53–0.78)	0.58 (0.50–0.66)	**0.87 (0.81–0.91)**	0.81 (0.64–0.92)
Specificity	Depression	0.62 (0.60–0.64)	0.59 (0.58–0.61)	0.59 (0.57–0.61)	0.59 (0.57–0.61)	0.58 (0.56–0.60)	0.59 (0.57–0.60)	0.58 (0.57–0.60)	0.58 (0.56–0.59)
	Anger	0.73 (0.71–0.74)	0.70 (0.68–0.71)	0.69 (0.68–0.71)	0.70 (0.68–0.71)	0.69 (0.67–0.70)	0.69 (0.68–0.71)	0.69 (0.68–0.71)	0.68 (0.67–0.70)
	Mania	0.74 (0.72–0.75)	0.73 (0.71–0.74)	0.72 (0.71–0.74)	0.73 (0.72–0.75)	0.72 (0.71–0.74)	0.73 (0.71–0.74)	0.73 (0.71–0.74)	0.72 (0.70–0.73)
	Anxiety	0.54 (0.52–0.56)	0.52 (0.50–0.53)	0.51 (0.49–0.53)	0.52 (0.50–0.53)	0.51 (0.49–0.52)	0.51 (0.49–0.53)	0.51 (0.49–0.52)	0.50 (0.48–0.52)
	Somatic symptoms	0.68 (0.66–0.69)	0.65 (0.64–0.67)	0.65 (0.64–0.67)	0.66 (0.64–0.67)	0.65 (0.63–0.66)	0.65 (0.63–0.66)	0.65 (0.63–0.66)	0.64 (0.62–0.66)
	Suicidal ideation	**0.92 (0.91–0.93)**	**0.92 (0.91–0.93)**	**0.92 (0.91–0.93)**	**0.92 (0.91–0.93)**	**0.92 (0.91–0.92)**	**0.92 (0.91–0.92)**	**0.91 (0.90–0.92)**	**0.91 (0.90–0.92)**
	Psychosis	**0.88 (0.86–0.89)**	**0.87 (0.86–0.88)**	**0.87 (0.85–0.88)**	**0.87 (0.86–0.88)**	**0.87 (0.86–0.88)**	**0.87 (0.86–0.88)**	**0.87 (0.86–0.88)**	**0.86 (0.85–0.87)**
	Sleep problems	**0.79 (0.77–0.80)**	0.77 (0.75–0.78)	0.77 (0.75–0.78)	**0.77 (0.76–0.78)**	0.76 (0.74–0.77)	0.76 (0.75–0.78)	0.77 (0.75–0.78)	0.75 (0.74–0.77)
	Memory	**0.86 (0.85–0.87)**	**0.85 (0.83–0.86)**	**0.84 (0.83–0.85)**	**0.85 (0.84–0.86)**	**0.84 (0.83–0.85)**	**0.84 (0.83–0.86)**	**0.84 (0.83–0.85)**	**0.83 (0.82–0.85)**

Domain								
Repetitive thoughts and behaviors	0.84 (0.83–0.85)	0.82 (0.80–0.83)	0.82 (0.80–0.83)	0.83 (0.81–0.84)	0.81 (0.80–0.83)	0.81 (0.80–0.83)	0.82 (0.80–0.83)	0.81 (0.79–0.82)
Dissociation	0.87 (0.86–0.88)	0.85 (0.84–0.86)	0.85 (0.84–0.86)	0.86 (0.84–0.87)	0.85 (0.84–0.86)	0.85 (0.84–0.86)	0.85 (0.84–0.86)	0.84 (0.83–0.85)
Personality functioning	0.74 (0.72–0.76)	0.72 (0.70–0.73)	0.72 (0.70–0.73)	0.72 (0.71–0.74)	0.71 (0.69–0.72)	0.71 (0.70–0.73)	0.72 (0.70–0.73)	0.70 (0.69–0.72)
Substance use	0.54 (0.52–0.56)	0.53 (0.52–0.55)	0.53 (0.51–0.55)	0.53 (0.52–0.55)	0.53 (0.51–0.55)	0.53 (0.51–0.55)	0.55 (0.53–0.57)	0.53 (0.51–0.55)

Sensitivity represents the ability of the test to designate an individual with the disorder as positive, and values greater than 0.75 are shown in bold if crossing the confidence interval. Specificity represents the ability of the test to designate an individual without the disorder as negative, and values greater than 0.75 are shown in bold if crossing the confidence interval.

Discussion

This is the first study to describe the screening properties of the DSM-XC for a broad spectrum of psychiatric diagnoses in a large community sample. Results suggest that screening for psychiatric conditions using simple and objective operational criteria increases the likelihood of detection or may help to rule out specific diagnoses. The results comprise three main findings: First, the DSM-XC domains are not diagnostic-specific, rather, they function as transdiagnostic screening tools. For example, a negative anxiety domain is useful in decreasing and a positive suicidal ideation domain is useful in increasing the probability of many internalizing disorders. Second, the screening utility of the DSM-XC is lower for externalizing disorders. Third, 'intensity' and 'frequency' framings as measured by the DSM-XC are only moderately correlated, and the 'frequency' framing exhibited a lower number of above-threshold domains and improved screening properties compared to 'intensity' framing.

While several screening tools exist for specific diagnosis, few address multiple diagnostic categories simultaneously. Despite limitations in externalizing disorder assessment, e.g. due to the lack of ADHD domain in the DSM-XC and the lack of a diagnostic interview for AUD in the present cohort, the DSM-XC performance for depression and anxiety is comparable to existing scales, justifying its potential applicability in community settings. The Patient Health Questionnaire-2 (PHQ-2) (Kroenke, Spitzer, & Williams, 2003) assesses similar symptoms to the depression domain of the DSM-XC, querying low mood and loss of interest over the past two weeks, despite differences in instructions and scoring scales. In the present study, we found screening properties of the DSM-XC 'OR' rule depression domain to detect MDD (LR + = 2.32, LR− = 0.08, sensitivity = 0.95, specificity = 0.59) to be similar to recent meta-analytic evidence examining the PHQ-2 (LR+ = 2.97, LR− = 0.26, sensitivity = 0.89, specificity = 0.75) (Mitchell et al., 2016). The Generalized Anxiety Disorder two questions scale (GAD-2) with a cut-off of three presents more informative properties (LR+ = 4.31, LR− = 0.25, sensitivity = 0.80, specificity = 0.81) (Plummer et al., 2016) compared to the DSM-XC anxiety domain to detect GAD (original 'OR' rule LR + 1.90, LR− = 0.24, sensitivity = 0.87, specificity = 0.54). The differences in performance between legacy measures with similar conceptual content and DSM-XC domains may be attributed to variations in item content, scale, and scoring methods (linear vs algorithm). For instance, the DSM-XC anxiety domain comprises three items, includes more symptoms per item and assesses different symptoms, such as panic and anxious avoidance, while the GAD-2 has only two items focusing on anxiety and worrying. Additionally, both PHQ-2 and GAD-2 use 4-point scales, while the DSM-XC is rated on a 5-point scale.

A lack of direct correspondence was observed across domains and psychiatric diagnoses. Moreover, some domains have a higher capability than others for screening, and some domains showed utility in screening for multiple disorders (transdiagnostic utility). For example, the depression domain showed a LR+ of 2.32 for detecting MDD, 2.11 for detecting GAD, and 1.94 for PTSD. Conversely, the anxiety domain showed significant LR− values for all internalizing disorders, suggesting its potential as a first screening step. This finding contrasts with the current understanding of the DSM-XC process of further inquiring an individual with disorder-specific questions if they screen positive in a domain. However, the DSM-XC did not show meaningful utility for externalizing disorders, except for a good likelihood ratio of a

Table 3. Likelihood ratios of DSM-XC domains to detect eight psychiatric disorders

Estimate, 95% CI	Domain	Internalizing disorders					Externalizing disorders		
		Generalized anxiety disorder	Social anxiety disorder	Major depressive disorder	Post-traumatic stress disorder	Bipolar disorder	Attention-deficit/ hyperactivity disorder	Alcohol use disorder	Antisocial personality disorder
Positive likelihood ratio	Depression	2.11 (1.97–2.25)	1.89 (1.72–2.06)	**2.32 (2.19–2.46)**	1.94 (1.77–2.12)	1.82 (1.58–2.11)	1.67 (1.50–1.87)	1.42 (1.26–1.60)	1.37 (1.04–1.82)
	Anger	**2.54 (2.33–2.77)**	1.99 (1.75–2.27)	**2.49 (2.21–2.80)**	**2.38 (2.12–2.66)**	**2.40 (2.06–2.80)**	1.93 (1.69–2.22)	1.60 (1.39–1.86)	1.40 (0.97–2.02)
	Mania	1.62 (1.43–1.85)	1.67 (1.41–1.98)	1.66 (1.33–2.05)	2.04 (1.75–2.37)	2.36 (1.95–2.87)	1.58 (1.31–1.89)	1.47 (1.24–1.75)	1.49 (1.01–2.20)
	Anxiety	1.90 (1.80–2.00)	1.78 (1.66–1.90)	1.95 (1.84–2.06)	1.94 (1.84–2.04)	1.92 (1.80–2.05)	1.52 (1.38–1.67)	1.28 (1.15–1.43)	1.39 (1.11–1.73)
	Somatic symptoms	2.06 (1.88–2.24)	1.73 (1.52–1.96)	2.16 (1.92–2.43)	2.10 (1.88–2.34)	2.22 (1.93–2.55)	1.54 (1.33–1.79)	1.39 (1.20–1.61)	1.39 (1.00–1.94)
	Suicidal ideation	**2.70 (2.14–3.42)**	**3.15 (2.39–4.15)**	**4.76 (3.64–6.23)**	**3.36 (2.54–4.46)**	**3.94 (2.71–5.72)**	2.20 (1.56–3.09)	1.75 (1.24–2.47)	1.25 (0.49–3.18)
	Psychosis	2.23 (1.85–2.69)	2.32 (1.83–2.94)	2.57 (1.95–3.40)	**3.08 (2.48–3.81)**	**4.02 (3.12–5.17)**	2.19 (1.70–2.83)	2.26 (1.80–2.84)	2.84 (1.87–4.31)
	Sleep problems	**2.53 (2.26–2.84)**	2.40 (2.08–2.78)	2.93 (2.54–3.37)	2.84 (2.49–3.24)	2.34 (1.86–2.94)	1.98 (1.66–2.36)	2.04 (1.75–2.38)	1.58 (1.04–2.39)
	Memory	**2.68 (2.30–3.13)**	2.72 (2.24–3.29)	2.45 (1.91–3.15)	**3.05 (2.53–3.67)**	**3.01 (2.29–3.95)**	2.37 (1.91–2.94)	1.79 (1.42–2.25)	1.51 (0.86–2.68)
	Repetitive thoughts and behaviors	**3.04 (2.67–3.46)**	2.50 (2.09–2.98)	**3.44 (2.92–4.05)**	**3.93 (3.46–4.48)**	**3.64 (3.03–4.38)**	2.27 (1.87–2.75)	2.03 (1.68–2.45)	2.15 (1.45–3.18)
	Dissociation	**3.06 (2.63–3.57)**	**2.66 (2.18–3.25)**	**3.95 (3.29–4.73)**	**3.38 (2.83–4.05)**	**3.94 (3.16–4.91)**	2.43 (1.95–3.02)	2.01 (1.61–2.50)	2.65 (1.79–3.93)
	Personality functioning	**2.41 (2.19–2.66)**	2.19 (1.93–2.49)	**3.04 (2.77–3.34)**	2.71 (2.44–3.01)	2.46 (2.08–2.90)	1.80 (1.54–2.10)	1.86 (1.62–2.14)	2.07 (1.59–2.70)
	Substance use	1.28 (1.17–1.41)	1.27 (1.12–1.44)	1.34 (1.15–1.56)	1.38 (1.22–1.56)	1.42 (1.18–1.70)	1.24 (1.08–1.42)	1.92 (1.80–2.05)	1.71 (1.45–2.02)
Negative likelihood ratio	Depression	**0.32 (0.26–0.39)**	0.39 (0.30–0.51)	**0.08 (0.03–0.19)**	**0.35 (0.26–0.48)**	0.40 (0.25–0.64)	0.53 (0.42–0.66)	0.70 (0.59–0.83)	0.72 (0.49–1.07)
	Anger	**0.42 (0.36–0.49)**	0.57 (0.47–0.68)	**0.34 (0.24–0.48)**	0.41 (0.32–0.52)	0.36 (0.23–0.56)	0.59 (0.49–0.71)	0.73 (0.64–0.84)	0.81 (0.61–1.09)
	Mania	0.78 (0.71–0.85)	0.75 (0.65–0.86)	0.75 (0.63–0.89)	0.62 (0.52–0.74)	0.48 (0.34–0.68)	0.78 (0.68–0.90)	0.82 (0.73–0.92)	0.81 (0.61–1.07)
	Anxiety	**0.24 (0.18–0.31)**	**0.27 (0.19–0.39)**	**0.09 (0.04–0.22)**	**0.12 (0.07–0.23)**	**0.10 (0.03–0.30)**	0.50 (0.38–0.65)	0.72 (0.60–0.87)	0.61 (0.37–1.00)
	Somatic symptoms	0.49 (0.43–0.57)	0.61 (0.51–0.73)	0.38 (0.27–0.53)	0.42 (0.33–0.55)	0.34 (0.21–0.54)	0.70 (0.59–0.84)	0.79 (0.69–0.90)	0.78 (0.56–1.08)
	Suicidal ideation	0.86 (0.82–0.91)	0.81 (0.75–0.89)	0.67 (0.58–0.78)	0.79 (0.72–0.87)	0.73 (0.61–0.87)	0.89 (0.83–0.96)	0.93 (0.88–0.99)	0.98 (0.87–1.10)
	Psychosis	0.83 (0.78–0.88)	0.80 (0.73–0.88)	0.76 (0.66–0.87)	0.70 (0.61–0.79)	0.54 (0.41–0.71)	0.82 (0.74–0.90)	0.81 (0.74–0.89)	0.71 (0.55–0.92)
	Sleep problems	0.58 (0.52–0.65)	0.58 (0.49–0.68)	0.41 (0.31–0.54)	0.45 (0.36–0.56)	0.57 (0.43–0.76)	0.70 (0.60–0.81)	0.68 (0.60–0.78)	0.81 (0.62–1.05)
	Memory	0.72 (0.67–0.78)	0.69 (0.61–0.78)	0.72 (0.62–0.85)	0.63 (0.55–0.73)	0.62 (0.48–0.79)	0.75 (0.66–0.84)	0.85 (0.78–0.93)	0.90 (0.74–1.09)

Repetitive thoughts and behaviors	0.61 (0.55–0.67)	0.67 (0.58–0.76)	0.45 (0.35–0.58)	**0.38 (0.30–0.48)**	0.39 (0.27–0.57)	0.71 (0.62–0.81)	0.77 (0.69–0.85)	0.72 (0.55–0.95)
Dissociation	0.69 (0.63–0.75)	0.71 (0.63–0.80)	0.49 (0.39–0.62)	0.60 (0.51–0.70)	0.47 (0.35–0.64)	0.75 (0.67–0.84)	0.82 (0.75–0.90)	0.69 (0.53–0.91)
Personality functioning	0.50 (0.44–0.58)	0.53 (0.44–0.64)	**0.20 (0.12–0.32)**	**0.34 (0.26–0.45)**	0.40 (0.27–0.60)	0.68 (0.58–0.79)	0.66 (0.57–0.76)	0.55 (0.37–0.83)
Substance use	0.76 (0.67–0.86)	0.76 (0.63–0.91)	0.70 (0.54–0.90)	0.66 (0.54–0.82)	0.63 (0.44–0.90)	0.79 (0.66–0.95)	**0.24 (0.17–0.35)**	0.37 (0.19–0.71)

LR values for positive tests greater than two (shown in bold if crossing the confidence interval) represent substantial changes in the probability of the target disorder, while LR values for negative tests less than 0.5 (also if crossing the interval) indicate considerable changes in the opposite direction.

negative screen in the substance use domain to rule out AUD. Due to identified shortcomings, specific recommended scales for alcohol and substance use should supplement primary care assessments using the DSM-XC. Further modifications in the tool are necessary to address externalizing symptoms, such as impulsivity and disruptive behaviors, aiming at increasing overall screening utility. These targeted improvements could broaden the DSM-XC's scope and solidify its place as a component of community screening strategies.

Strong connections observed in certain domains' screening capabilities likely reflect the interconnectedness of the disorders they captured. The presence of clinical features that are not represented in the diagnostic criteria but are frequent among individuals with certain mental disorders is a potential explanation for some domains demonstrating screening utility for multiple disorders. For instance, somatic symptoms, while not part of diagnostic criteria, may be more commonly experienced among individuals with GAD and MDD, while anxious distress is a notable feature in both bipolar and major depressive disorders (Bartoli et al., 2024). Another contributing factor is the conceptual similarity and symptom overlap across various diagnostic criteria, with insomnia and irritable mood being among the top non-specific symptoms (Forbes et al., 2024). In our study, sleep problems and anger, for example, demonstrated significant positive likelihood ratios across anxiety, mood, and post-traumatic stress disorder categories. Comorbidity and common temperamental risk factors can contribute to this phenomenon, as individuals with high levels of negative affectivity (neuroticism) are likely to endorse multiple internalizing symptom domains in a non-specific manner due to being oriented towards negative affectivity (Griffith et al., 2010). Additionally, substantial evidence supports the idea that common neural and genetic mechanisms contribute to the association between these disorders (Jami et al., 2022; Xie et al., 2023).

We also examined the contributions of framing questions as intensity or frequency of symptoms, as the questionnaire was revised based on participant feedback regarding the challenge of understanding double-barreled questions. The DSM-XC showed a moderate correlation between framings, with intensity-framed questions leading to a significantly higher proportion of positive screens for most domains compared to frequency-framed questions. For some domains, such as depression, anger, and memory, framing questions as 'intensity' approximately doubled the number of individuals that would have been needed to complete Level 2 measures, compared with 'frequency'. Thus, using frequency-framed questions may be a more resource-efficient approach.

Framings also differed in screening properties, with frequency-framed questions presenting generally higher specificity but equivalent sensitivity to detect their corresponding specific disorders. Differences in LR+ favoring the use of the 'frequency' framing were found, with minimal compromise in LR− values. The value of likelihood ratios for clinical decision-making is significant (Grimes & Schulz, 2005). For example, a positive result in the anxiety domain, with a likelihood ratio of 1.95, translates to a shift from the pre-test probability of GAD from 10.5% to a post-test probability of 18.3%. However, by employing the 'frequency' framing, with a LR+ value of 3.11, the post-test probability substantially increases to 26.8% for GAD. As the conversion from pre-test to post-test probability involves a transformation into odds, we have developed an accessible tool, available at https://pacheco-jpg.shinyapps.io/dsm_xc_shiny_app/ that enables quick estimation of the shift from pre- to post-test probability for the tested disorders within the DSM-XC domains.
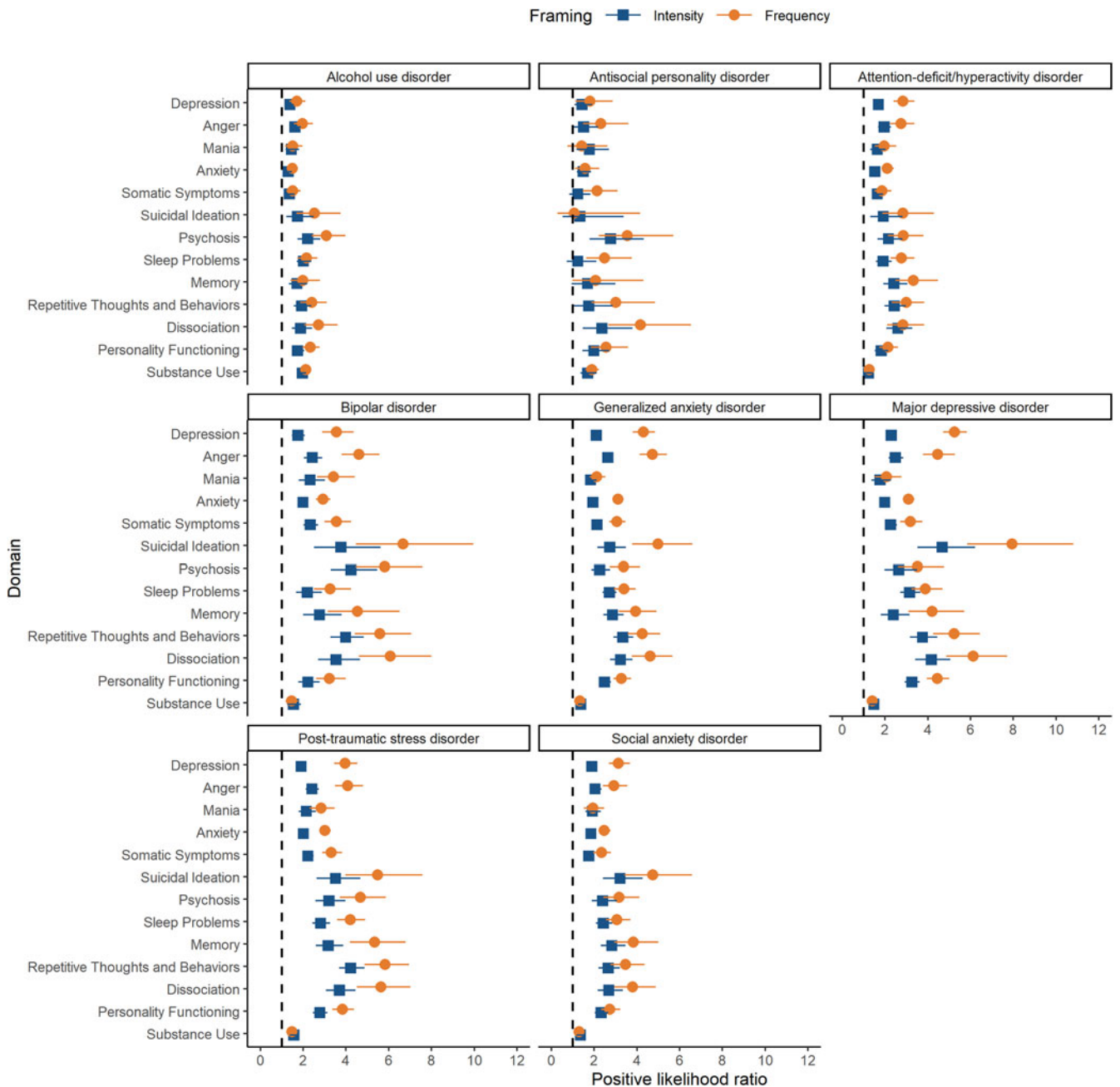
Figure 1. Positive likelihood ratio (LR+) across DSM-XC domains for psychiatric diagnoses using intensity and frequency framings. **Legend:** Intensity (in blue) and frequency (in yellow) represent how the instrument would perform if questions were asked with separate frames. DSM-XC, DSM-5-TR Level 1 Cross-Cutting Symptom Measure.

Apart from establishing improved screening utility for the DSM-XC, defining standards for adverbial framing of questions can reduce the substantial heterogeneity in mental health assessment (Newson, Hunter, & Thiagarajan, 2020). Previous work suggests higher reliability, criterion validity (Krishnakumar et al., 2021), and stability over time (Krabbe & Forkmann, 2014) for frequency in comparison to intensity, which could explain the higher accuracy of the frequency framing that we found in the present study.

This study has three main strengths. First, we examined a community-based sample, which allows for real-world estimates of the screening properties of the DSM-XC. Second, we

investigated a broad spectrum of psychiatric diagnoses, with examples of both internalizing and externalizing disorders. Third, we removed a previous restriction imposed by double-barreled phrasing, which allowed for examining the contributions of intensity and frequency framings. However, interpreting these results requires considering several limitations. First, results should not be necessarily extrapolated for other age groups as the sample was limited to the age of 22. Second, although the presence of most disorders was determined by a qualified clinical psychologist with a structured interview, we used the AUDIT to evaluate AUD. Third, further research is needed to directly compare the implemented 'OR' rule with the original double-barreled
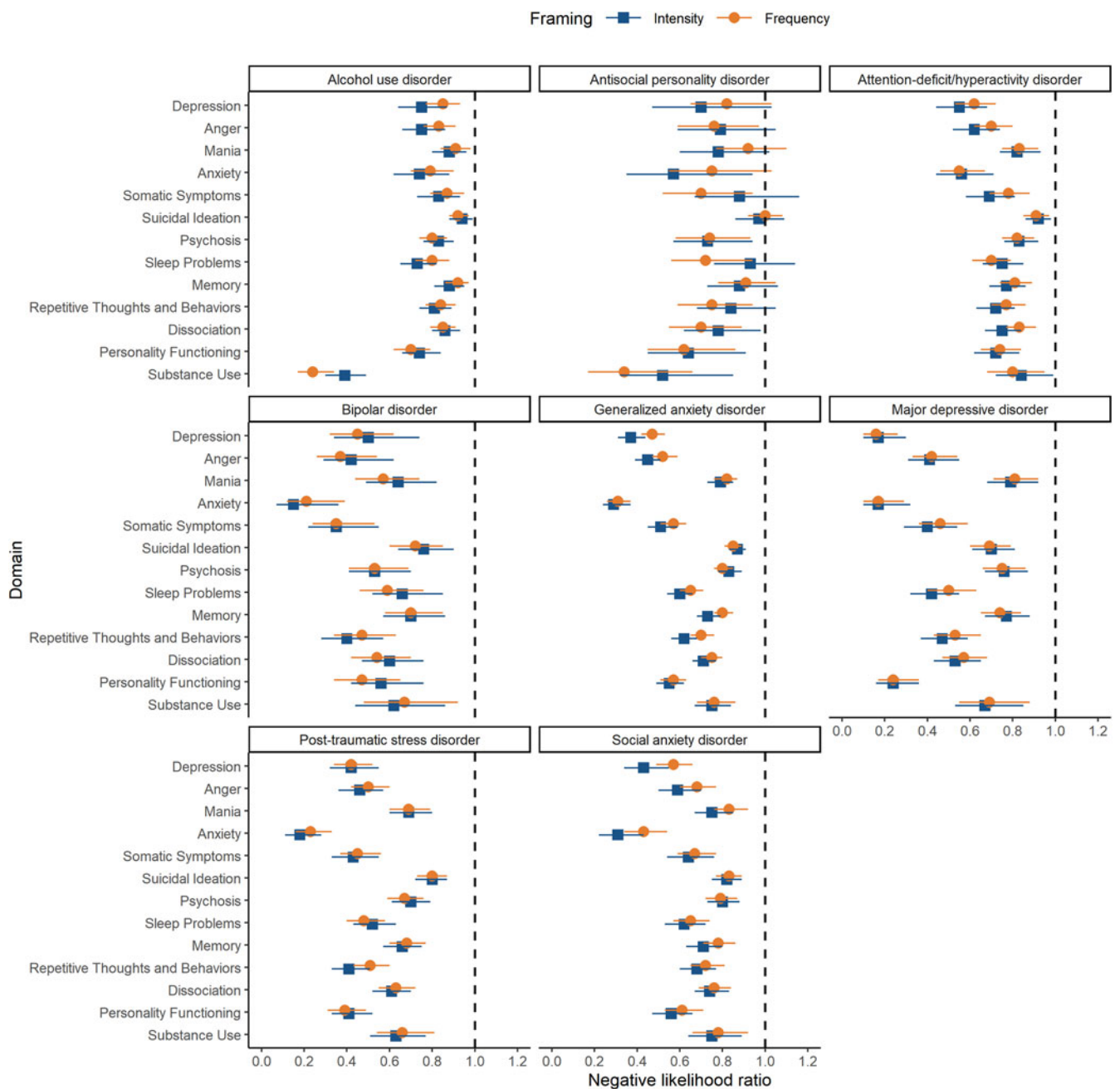
Figure 2. Negative likelihood ratio (LR+) across DSM-XC domains for psychiatric diagnoses using intensity and frequency framings. **Legend:** Intensity (in blue) and frequency (in yellow) represent how the instrument would perform if questions were asked with separate frames. DSM-XC, DSM-5-TR Level 1 Cross-Cutting Symptom Measure.

scale. The 'OR' rule allowed us to characterize the intended meaning of the DSM-XC, which is an important as it addresses the cognitively demanding phrasing of the original scale. A notable concern with the implemented approach is that the order in that questions were presented could have affected how individuals rated each item. As the 'intensity' framing was always presented first, this piece of information could have influenced how participants answered the following 'frequency' question, introducing anchoring bias. However, anchoring bias tends to make results look more similar and we found only moderate correlations between framings. Additionally, the frequency framing, presented after intensity, showed better screening properties, and logistic

regression analyses of specific framing subgroups suggest the frequency framing captures relevant information of specific corresponding disorders on more occasions than the intensity framing alone. This suggests that anchoring bias did not occur as frequency framing is not conditioned to intensity framing.

The present results demonstrated that the DSM-XC is useful in screening multiple domains of psychopathology, increasing or decreasing the likelihood of detection of specific diagnoses with simple objective questions. Due to the low direct correspondence between domain and diagnosis, more than one Level 2 Cross-Cutting Symptom Measures could apply after a positive screening in some domains. We also found that adverbial

framings have an impact on the DSM-XC's screening performance. Framing questions as 'frequency' may be more resource-efficient and could lead to improved detection of positive cases. We recommend that these findings should be replicated in other samples to allow implementation in routine practice and that future examination of scoring thresholds should be done to further optimize the DSM-XC as a screening tool for mental health conditions.

References

Altman, D. G., & Bland, J. M. (1994). Statistics notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ*, *308*(6943), 1552–1552. doi:10.1136/bmj.308.6943.1552

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Publishing.

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

Amorim, P. (2000). Mini international neuropsychiatric interview (MINI): Mini international neuropsychiatric interview (MINI): Validation of a short structured diagnostic psychiatric interview. *Brazilian Journal of Psychiatry*, *22*(3), 106–115. doi:10.1590/S1516-44462000000300003

Bartoli, F., Bachi, B., Callovini, T., Palpella, D., Piacenti, S., Morreale, M., … Carrà, G. (2024). Anxious distress in people with major depressive episodes: A cross-sectional analysis of clinical correlates. *CNS Spectrums*, *29*(1), 49–53. doi:10.1017/S1092852923002377

Bastiaens, L., & Galus, J. (2018). The DSM-5 self-rated Level 1 cross-cutting symptom measure as a screening tool. *Psychiatric Quarterly*, *89*(1), 111–115. https://doi.org/10.1007/s11126-017-9518-7

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the hospital anxiety and depression scale. *Journal of Psychosomatic Research*, *52*(2), 69–77. doi:10.1016/S0022-3999(01)00296-3

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., … Cohen, J. F. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, *351*, h5527. doi:10.1136/bmj.h5527

Collett, D. (2002). *Modelling binary data* (2nd ed.). New York, NY: Chapman and Hall/CRC. doi:10.1201/b16654

Doss, R. A., & Lowmaster, S. E. (2022). Validation of the DSM-5 Level 1 cross-cutting symptom measure in a community sample. *Psychiatry Research*, *318*, 114935. doi:10.1016/j.psychres.2022.114935

Eaton, N. R., Bringmann, L. F., Elmer, T., Fried, E. I., Forbes, M. K., Greene, A. L., … Waszczuk, M. A. (2023). A review of approaches and models in psychopathology conceptualization research. *Nature Reviews Psychology*, *2*, 622–636. doi:10.1038/s44159-023-00218-4

Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., … Dras, M. (2024). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychological Medicine*, *54*(5), 886–894. doi:10.1017/S0033291723002544

Forbes, M. K., Tackett, J. L., Markon, K. E., & Krueger, R. F. (2016). Beyond comorbidity: Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and Psychopathology*, *28*(4pt1), 971–986. doi:10.1017/S0954579416000651

Gibbons, A. B., Farmer, C., Shaw, J. S., & Chung, J. Y. (2023). Examining the factor structure of the DSM-5 Level 1 cross-cutting symptom measure. *International Journal of Methods in Psychiatric Research*, *32*(2), e1953. doi:10.1002/mpr.1953

Gonçalves, H., Wehrmeister, F. C., Assunção, M. C. F., Tovo-Rodrigues, L., de Oliveira, I. O., Murray, J., … Menezes, A. M. B. (2018). Cohort profile update: The 1993 Pelotas (Brazil) birth cohort follow-up at 22 years. *International Journal of Epidemiology*, *47*(5), 1389–1390e. doi:10.1093/ije/dyx249

Gong, J., Zhang, Y., Yang, Z., Huang, Y., Feng, J., & Zhang, W. (2013). The framing effect in medical decision-making: A review of the literature. *Psychology, Health & Medicine*, *18*(6), 645–653. doi:10.1080/13548506.2013.766352

Griffith, J. W., Zinbarg, R. E., Craske, M. G., Mineka, S., Rose, R. D., Waters, A. M., & Sutton, J. M. (2010). Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine*, *40*(7), 1125–1136. doi:10.1017/S0033291709991449

Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet*, *365*(9469), 1500–1505. doi:10.1016/S0140-6736(05)66422-7

Gryczynski, J., Kelly, S. M., Mitchell, S. G., Kirk, A., O'Grady, K. E., & Schwartz, R. P. (2015). Validation and performance of the alcohol, smoking and substance involvement screening test (ASSIST) among adolescent primary care patients. *Addiction*, *110*(2), 240–247. doi:10.1111/add.12767

Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. *JAMA*, *271*(9), 703–707. doi:10.1001/jama.271.9.703

Jami, E. S., Hammerschlag, A. R., Ip, H. F., Allegrini, A. G., Benyamin, B., Border, R., … Middeldorp, C. M. (2022). Genome-wide association meta-analysis of childhood and adolescent internalizing symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(7), 934–945. doi:10.1016/j.jaac.2021.11.035

Kassambara, A. (2021). *Package 'rstatix' – pipe-friendly framework for basic statistical tests.* https://CRAN.R-project.org/package=rstatix

Krabbe, J., & Forkmann, T. (2014). Frequency vs. intensity: Framing effects on patients' use of verbal rating scale anchors. *Comprehensive Psychiatry*, *55*(8), 1928–1936. doi:10.1016/j.comppsych.2014.06.010

Krishnakumar, A., Scopel Hoffmann, M., Schoeller, F., Clucas, J. C., Son, J., Müller-Naendrup, L., … Klein, A. (2021). Framing mental health questions using context better predicts disability than does frequency. *PsyArXiv* [Preprint]. doi:10.31234/osf.io/jm9kc

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292. doi:10.1097/01.MLR.0000093487.78664.3C

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. doi:10.1006/obhd.1998.2804

Löwe, B., Kroenke, K., & Gräfe, K. (2005). Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *Journal of Psychosomatic Research*, *58*(2), 163–171. doi:10.1016/j.jpsychores.2004.09.006

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., … Brähler, E. (2010). A 4-item measure of depression and anxiety:

Validation and standardization of the patient health questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, 122 (1–2), 86–95. doi:10.1016/j.jad.2009.06.019

Mahoney, M. R., Farmer, C., Sinclair, S., Sung, S., Dehaut, K., & Chung, J. Y. (2020). Utilization of the DSM-5 self-rated Level 1 cross-cutting symptom measure-adult to screen healthy volunteers for research studies. *Psychiatry Research*, 286, 112822. doi:10.1016/j.psychres.2020.112822

Manfro, P. H., Belem da Silva, C. T., Anselmi, L., Barros, F., Eaton, W. W., Gonçalves, H., … Kieling, C. (2021). Depression in a youth population-based sample from Brazil: Prevalence and symptom structure. *Journal of Affective Disorders*, 292, 633–641. doi:10.1016/j.jad.2021.05.073

Matte, B., Anselmi, L., Salum, G. A., Kieling, C., Gonçalves, H., Menezes, A., … Rohde, L. A. (2015). ADHD in DSM-5: A field trial in a large, representative sample of 18- to 19-year-old adults. *Psychological Medicine*, 45(2), 361–373. doi:10.1017/S0033291714001470

McNeely, J., Wu, L.-T., Subramaniam, G., Sharma, G., Cathers, L. A., Svikis, D., … Schwartz, R. P. (2016). Performance of the tobacco, alcohol, prescription medication, and other substance use (TAPS) tool for substance use screening in primary care patients. *Annals of Internal Medicine*, 165(10), 690. doi:10.7326/M16-0317

Means-Christensen, A. J., Arnau, R. C., Tonidandel, A. M., Bramson, R., & Meagher, M. W. (2005). An efficient method of identifying major depression and panic disorder in primary care. *Journal of Behavioral Medicine*, 28(6), 565–572. doi:10.1007/s10865-005-9023-6

Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., Hallak, J. E., Trzesniak, C., de Azevedo Marques, J. M., … Crippa, J. A. (2010). Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcoholism, clinical and experimental research*, 34 (8), 1417–1424. doi:10.1111/j.1530-0277.2010.01225.x

Mitchell, A. J., Yadegarfar, M., Gill, J., & Stubbs, B. (2016). Case finding and screening clinical utility of the patient health questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic meta-analysis of 40 studies. *BJPsych Open*, 2(2), 127–138. doi:10.1192/bjpo.bp.115.001685

Mulvaney-Day, N., Marshall, T., Downey Piscopo, K., Korsen, N., Lynch, S., Karnell, L. H., … Ghose, S. S. (2018). Screening for behavioral health conditions in primary care settings: A systematic review of the literature. *Journal of General Internal Medicine*, 33(3), 335–346. doi:10.1007/s11606-017-4181-0

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., … Thombs, B. D. (2021). Accuracy of the patient health questionnaire-9 for screening to detect major depression: Updated systematic review and individual participant data meta-analysis. *BMJ*, 375, n2183. doi:10.1136/bmj.n2183

Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The heterogeneity of mental health assessment. *Frontiers in Psychiatry*, 11, 76. doi:10.3389/fpsyt.2020.00076

Plana-Ripoll, O., Pedersen, C. B., Holtz, Y., Benros, M. E., Dalsgaard, S., de Jonge, P., … McGrath, J. J. (2019). Exploring comorbidity within mental disorders among a Danish national population. *JAMA Psychiatry*, 76(3), 259. https://doi.org/10.1001/jamapsychiatry.2018.3658

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. *General Hospital Psychiatry*, 39, 24–31. doi:10.1016/j.genhosppsych.2015.11.005

Power, M., Fell, G., & Wright, M. (2013). Principles for high-quality, high-value testing. *Evidence Based Medicine*, 18(1), 5–10. doi:10.1136/eb-2012-100645

Simel, D. L., Samsa, G. P., & Matchar, D. B. (1991). Likelihood ratios with confidence: Sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*, 44(8), 763–770. doi:10.1016/0895-4356(91)90128-V

Smith, P. C., Schmidt, S. M., Allensworth-Davies, D., & Saitz, R. (2009). Primary care validation of a single-question alcohol screening test. *Journal of General Internal Medicine*, 24(7), 783–788. doi:10.1007/s11606-009-0928-6

Stevenson, M., Sergeant E., Heuer C., Nunes T., Heuer C., Marshall J., … Rabiee A. (2020). *Package 'epiR' – tools for the analysis of epidemiological data*. https://CRAN.R-project.org/package=epiR

US Preventive Services Task Force, Barry, M. J., Nicholson, W. K., Silverstein, M., Chelmow, D., Coker, T. R., … Wong, J. B. (2023b). Screening for depression and suicide risk in adults: US preventive services task force recommendation statement. *JAMA*, 329(23), 2057–2067. doi:10.1001/jama.2023.9297

US Preventive Services Task Force, Barry, M. J., Nicholson, W. K., Silverstein, M., Coker, T. R., Davidson, K. W., … Wong, J. B. (2023a). Screening for anxiety disorders in adults: US preventive services task force recommendation statement. *JAMA*, 329(24), 2163–2170. doi:10.1001/jama.2023.9301

US Preventive Services Task Force, Curry, S. J., Krist, A. H., Owens, D. K., Barry, M. J., Caughey, A. B., … Wong, J. B. (2018). Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US preventive services task force recommendation statement. *JAMA*, 320(18), 1899–1909. doi:10.1001/jama.2018.16789

US Preventive Services Task Force, Krist, A. H., Davidson, K. W., Mangione, C. M., Barry, M. J., Cabana, M., … Wong, J. B. (2020). Screening for unhealthy drug use: US preventive services task force recommendation statement. *JAMA*, 323(22), 2301–2309. doi:10.1001/jama.2020.8020

Victora, C. G., Araújo, C. L. P., Menezes, A. M. B., Hallal, P. C., Vieira, M. d. F., Neutzling, M. B., … Barros, F. C. (2006). Methodological aspects of the 1993 Pelotas (Brazil) birth cohort study. *Revista de Saúde Pública*, 40(1), 39–46. doi:10.1590/S0034-89102006000100008

Victora, C. G., Hallal, P. C., Araujo, C. L., Menezes, A. M., Wells, J. C., & Barros, F. C. (2008). Cohort profile: The 1993 Pelotas (Brazil) birth cohort study. *International Journal of Epidemiology*, 37(4), 704–709. doi:10.1093/ije/dym177

Vilagut, G., Forero, C. G., Barbaglia, G., & Alonso, J. (2016). Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): A systematic review with meta-analysis. *PLoS One*, 11(5), e0155431. doi:10.1371/journal.pone.0155431

Waszczuk, M. A., Zimmerman, M., Ruggero, C., Li, K., MacNamara, A., Weinberg, A., … Kotov, R. (2017). What do clinicians treat: Diagnoses or symptoms? The incremental validity of a symptom-based, dimensional characterization of emotional disorders in predicting medication prescription patterns. *Comprehensive Psychiatry*, 79, 80–88. doi:10.1016/j.comppsych.2017.04.004

Xie, C., Xiang, S., Shen, C., Peng, X., Kang, J., Li, Y., … ZIB Consortium. (2023). A shared neural basis underlying psychiatric comorbidity. *Nature Medicine*, 29(5), 1232–1242. doi:10.1038/s41591-023-02317-4