



The Heterogeneity of Mental Health Assessment

Jennifer J. Newson*, Daniel Hunter and Tara C. Thiagarajan

Sapien Labs, Arlington, VA, United States

OPEN ACCESS

Edited by:

Amitai Abramovitch,
Texas State University,
United States

Reviewed by:

Caleb Lack,
University of Central Oklahoma,
United States
Joseph McGuire,
Johns Hopkins University,
United States
Joseph Etherton,
Texas State University,
United States

*Correspondence:

Jennifer J. Newson
jennifer@sapienlabs.org

Specialty section:

This article was submitted to
Public Mental Health,
a section of the journal
Frontiers in Psychiatry

Received: 09 October 2019

Accepted: 30 January 2020

Published: 27 February 2020

Citation:

Newson JJ, Hunter D and
Thiagarajan TC (2020) The
Heterogeneity of Mental
Health Assessment.
Front. Psychiatry 11:76.
doi: 10.3389/fpsy.2020.00076

Across the landscape of mental health research and diagnosis, there is a diverse range of questionnaires and interviews available for use by clinicians and researchers to determine patient treatment plans or investigate internal and external etiologies. Although individually, these tools have each been assessed for their validity and reliability, there is little research examining the consistency between them in terms of what symptoms they assess, and how they assess those symptoms. Here, we provide an analysis of 126 different questionnaires and interviews commonly used to diagnose and screen for 10 different disorder types including depression, anxiety, obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), attention deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), addiction, bipolar disorder, eating disorder, and schizophrenia, as well as comparator questionnaires and interviews that offer an all-in-one cross-disorder assessment of mental health. We demonstrate substantial inconsistency in the inclusion and emphasis of symptoms assessed within disorders as well as considerable symptom overlap across disorder-specific tools. Within the same disorder, similarity scores across assessment tools ranged from 29% for assessment of bipolar disorder to a maximum of 58% for OCD. Furthermore, when looking across disorders, 60% of symptoms were assessed in at least half of all disorders illustrating the extensive overlap in symptom profiles between disorder-specific assessment tools. Biases in assessment toward emotional, cognitive, physical or behavioral symptoms were also observed, further adding to the heterogeneity across assessments. Analysis of other characteristics such as the time period over which symptoms were assessed, as well as whether there was a focus toward frequency, severity or duration of symptoms also varied substantially across assessment tools. The consequence of this inconsistent and heterogeneous assessment landscape is that it hinders clinical diagnosis and treatment and frustrates understanding of the social, environmental, and biological factors that contribute to mental health symptoms and disorders. Altogether, it underscores the need for standardized assessment tools that are more disorder agnostic and span the full spectrum of mental health symptoms to aid the understanding of underlying etiologies and the discovery of new treatments for psychiatric dysfunction.

Keywords: mental health, psychiatric, diagnosis, transdiagnostic, questionnaires, assessment, cross-disorder

INTRODUCTION

Across clinical and research domains, mental health assessment and diagnosis are carried out using interviews and questionnaires that determine the presence, severity, frequency, and duration of a broad range of psychiatric symptoms. The question content of these assessment tools is often based on classification systems such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (1) or International Classification of Diseases (ICD-11) (2) where pre-defined patterns of symptom criteria have been grouped together and designated as specific mental health disorders. Their design ranges from more open-ended clinician-led interviews typically used to make a formal psychiatric diagnosis (e.g., SCID) (3), to more quantitatively designed auxiliary questionnaires (e.g., PHQ-9) (4) that provide multidimensional assessments of symptom experience and severity to support diagnosis and treatment evaluation in clinical practice, and that are used to investigate underlying etiologies and treatment effectiveness in clinical trials and academic research studies.

A top down perspective of this landscape of mental health assessment reveals a huge range of interviews and questionnaires available for use. This diversity of choice means there is no shortage of options when searching for assessment tools for clinical use or to suit the needs of a clinical research study. However, this diversity can also make it a real challenge to decide which questionnaire(s) or interview(s) to select for clinical diagnosis or evaluation. For example, there have been more than 280 different questionnaires developed over the last century to assess symptoms of depression (5) which differ in terms of which iteration of the DSM they align to, the degree to which they consider co-morbid symptoms, whether they are computer based or paper based, and whether they are self-rated, parent-rated or clinician-led. Knowing which questionnaire to choose to obtain a suitable assessment of an individual's mental health is therefore not always a straightforward exercise for even the most experienced researcher or clinician.

The significance of having so many assessment options is two-fold. Firstly, within the clinic, screening questionnaires and interviews allow clinicians to build up a picture of the psychological problems and concerns faced by their patients so they can determine a diagnosis that will then determine a treatment regime. However, when different clinicians use different assessment tools during patient evaluations, it has the potential to introduce variability and inconsistency in diagnosis (6). Furthermore, as many assessment and diagnostic tools are disorder specific, a patient's experience rarely fits neatly within these theoretically defined boundaries, and therefore the choice of tool can potentially bias diagnosis. In addition, the onset and trajectory of a mental health disorder is often impacted by numerous social and environmental factors. If clinicians use assessment tools that differ by the degree to which they ask about these factors (e.g., triggers), especially for disorders where external events are known to have a considerable impact (e.g., depression), then it may provide a varying or incomplete picture of these contributing elements, which in turn could hinder the delivery of appropriate patient treatment plans, or provide

conflicting evidence in relation to mental health recommendations and policies.

Secondly, across the landscape of mental health research there is a drive to better understand underlying etiologies to help deliver new treatment opportunities. One avenue is the development of physiological biomarkers to aid diagnosis of mental health disorders and guide discovery of pharmacological intervention (7–12), although see (13, 14). These studies often aim to reveal correlations between neurobiological changes in a patient's brain (e.g., using tools such as fMRI and EEG) and the presence or severity of a mental health diagnosis. However, when these studies use different questionnaires and interviews to review the characteristics and severity of a patient's thoughts, feelings, and actions, it has the potential to introduce considerable variability that hinders comparisons across studies and reproducibility. Similarly, when researchers use different questionnaires or interviews in the assessment of symptom severity in studies evaluating the effectiveness of new or existing pharmacological, cognitive and/or behavioral therapies, or when trying to understand psychopathological mechanisms or the phenomenology of a disorder, it again has the potential to introduce variability and impede attempts to demonstrate the validity and reliability of results across studies.

Despite the large variety of assessment and diagnostic tools, there is relatively little research that has provided an in-depth analysis of the consistency of symptoms assessed across questionnaires both within and between different mental health disorders [although see (15) for a within disorder perspective of depression]. For the purposes of this paper, the term consistency denotes the degree to which different assessment tools designed for the same disorder assess the same set of symptoms, and whether they assess those symptoms using similar questioning characteristics. By identifying the level of consistency, or inconsistency, across assessment tools it allows us to determine the true extent of the potential problems noted above, and also allows researchers and clinicians to see how their favored choice of assessment tool(s) could be influencing or biasing their assessment or results. For example, if one clinician preferably uses a tool that predominantly assesses behaviors and actions, while another uses a tool (developed for the same disorder) that focuses on emotional difficulties, then the clinicians may end up creating different clinical impressions of the patients in front of them.

To explore this, we carried out an analysis of 126 questionnaires and interviews used to diagnose and screen for 10 different disorder types, including depression, anxiety, obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), attention deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), addiction (drug and alcohol), bipolar disorder, eating disorder, and schizophrenia, as well as comparator questionnaires and interviews that offer an all-in-one cross-disorder assessment of mental health. These cross-disorder tools were included to provide a perspective on the breadth of symptoms assessed by each tool to reveal which ones covered the widest spectrum of mental health symptoms, as well as to determine the consistency of symptoms assessed by these tools.

As this analysis was conducted purely on the written content of the questions within an assessment or diagnostic tool, rather than on responses made by any participant group, our aim was not to determine the reliability or consistency of responses to individual questionnaires, nor was it to cluster symptoms according to patient responses to these questionnaires. Instead, our objective was to investigate the range and diversity of symptoms assessed across these tools as well as the way those symptoms were assessed, to explore any potential inconsistencies in the landscape of mental health assessment. In particular, we hypothesized that screening tools within individual disorders would be highly consistent across core sets of symptoms, revealing standardized diagnostic criteria. On the other hand, we hypothesized that while there would also be some overlap in the symptoms assessed across disorders, individual disorders would be relatively distinct in symptom profile. In contrast to this *a priori* hypothesis we found a highly inconsistent and heterogeneous landscape within and across disorders. The specific set of symptoms assessed within disorders, as well as the degree to which assessments focused on cognitive, emotional, physical or behavioral symptoms, and the extent to which they asked about symptom triggers or consequences, varied dramatically as did assessments of symptom time frame, and whether they focused on the presence, frequency, duration or timing of symptoms. Furthermore, most symptoms were very broadly distributed across multiple disorders rather than being disorder specific. We also hypothesized that cross-disorder screening tools would cover a wide breadth of symptoms that spanned symptoms across all major mental health disorders with a high degree of similarity. However, we found the opposite, where cross-disorder tools were generally incomplete and differed in the breadth of symptoms they asked about as well as the degree to which they focused on emotional, behavioral, physical, and cognitive symptoms.

MATERIALS AND METHODS

We conducted an analysis of clinical interviews and screening questionnaires commonly used within a clinical or research domain. The analysis focused on 10 psychiatric disorders including depression, anxiety, OCD, PTSD ADHD, ASD, addiction, bipolar disorder, eating disorder and schizophrenia, covering questionnaires and interviews developed for adult, and pediatric populations. These disorders were selected based on a review of the disorders included in the DSM clinical interview (SCID-CV) (3). In addition, ASD and eating disorder were included due to both their prevalence and their broad public and scientific interest. Also included are a number of commonly used questionnaires and interviews which were not specific to any one disorder, but which instead took a cross-disorder approach to mental health assessment.

Coding and Analysis

The study was conducted in three steps. First, we identified and selected a comprehensive set of psychiatric questionnaires and clinical interviews that were commonly cited in the scientific literature. Next, we systematically categorized the individual questions included in each of the selected questionnaires by

symptoms and themes. Finally, we compared the symptoms assessed across questionnaires and interviews, both within individual disorder types, and from a cross-disorder perspective, computing similarity scores across all.

Questionnaires Identification and Selection

To select the questionnaires, we conducted several different searches and reviews of the literature. Firstly, we conducted a search of PubMed¹ (November 2019) using combinations of the following keywords in the title “questionnaire OR interview OR rating OR scale OR inventory OR instrument OR measure” alongside the key terms for each disorder of interest. This revealed a total of 10,318 search entries, from which the names of 929 different assessment tools were extracted. Secondly, we collated together a set of meta-analyses and reviews which specifically compared different assessment methodologies for each of the 10 disorders and identified the common questionnaires and interviews reviewed in those publications. Thirdly, we reviewed a number of websites and publishers which included lists of psychiatric assessment tools. Finally, we went through a large-scale cross-disorder review which covered 9 out of the 10 psychiatric disorders included in this review (16). Altogether, these searches revealed a large number of different screening questionnaires and clinical interviews.

To narrow down this larger group of questionnaires and interviews, we first identified the primary citation reference for each assessment tool and then identified the number of Google Scholar citations for each as a measure of use in clinical and academic research. We then excluded assessment tools based on the following criteria: (i) those that specifically covered multiple disorders (e.g., assessed anxiety and depression together); (ii) those that only assessed a specific subset of symptoms associated with a disorder (e.g., only cognitive aspects); (iii) those that only targeted a specific clinical group (e.g., stroke patients); (iv) those that had been used to assess a disorder but weren't specific to that disorder (e.g., when an OCD tool is used to assess symptoms of schizophrenia); (v) those which were not specific to any one disorder (e.g., assessing medication compliance). We then selected the questionnaires and interviews with higher citation numbers, while also taking into account their publication date. Although we acknowledge that the citation numbers provided by Google Scholar are often higher than those provided from other sources (e.g., Web of Science) we considered this approach to be acceptable when comparing between references. Where accurate citation numbers were not readily available from Google Scholar (e.g., if the questionnaire was reported in a manual or book) then we made a judgement according to how frequently the tool was mentioned/used across the studies we reviewed as to whether it should be included within our final analysis. The final list of questionnaires and interviews we have included is by no means exhaustive but aims to include as many of the most common tools used within clinical and research settings as possible.

Altogether 126 questionnaires and clinical interviews published between 1959 and 2018 were selected for inclusion in the final analysis (**Table 1, Supplementary Table 1**) that

¹<https://www.ncbi.nlm.nih.gov/pubmed/>

TABLE 1 | Overview of Questionnaires and Interviews included in the analysis.

Disorder	Number of Questionnaires	Number of questions reviewed	List of Assessment Tools Included*
Depression	19	369	<i>Adult:</i> APA-Dep-A ¹ , BDI-II (17), CES-D (18), CESDR (19), EPDS (20), GDS-LF (21), GDS-SF (22), HAMD (23), IDS (24), MADRS (25), PHQ9 (4), QIDS (26), ZDS (27). <i>Pediatric:</i> APA-Dep-C ¹ , CDI2 (28, 29), CESDC (30), MFQ (31), MFQS (31), RADS2 (32).
Anxiety	13**	483	<i>Adult:</i> APA-Anx-A ^{1**} , BAI (33), GAD7 (34), HAMA (35), LSAS (36, 37), SPIN (38), STAI (39), ZAS (40). <i>Pediatric:</i> APA-Anx-C ^{1**} , MASC (41, 42), RCMAS (43), SCARED (44, 45), SCAS (46).
PTSD	9	376	<i>Adult:</i> APA-PTSD-A ¹ , CAPS-5 (47), PC-PTSD-5 (48), PCL-5 (49), PSS-SR5 (50), SPRINT (51). <i>Pediatric:</i> APA-PTSD-C ¹ , CAPS-5-CV (52), CPSS-V (53).
Bipolar/Mania	5	90	<i>Adult:</i> ASRMS (54), HCL32 (55), ISS (56), MDQ (57), YMS (58).
OCD	8	330	<i>Adult:</i> DOCS (59), FOCI™ (60), OCI-R (61), PI-WSUR (62), VOCl (63), Y-BOCS™ (64, 65). <i>Pediatric:</i> CY-BOCS™ (66), OCI-CV (67).
Addiction	13	319	<i>Adult:</i> ADS (68, 69), ASI-5 (70), ASSIST-3 (71), AUDIT (72), CAGE (73), DAST-10 (74), DAST-20 (74), DUDIT (75), MAST (76), OCDS (77), SMAST (78), TWEAK (79). <i>Pediatric:</i> CRAFFT (80).
ASD	22	1213	<i>Adult:</i> AAA (81), AQ-A-10 (82), AQ-A (83), BAPQ (84), EQ-A (85), SQ-A (86). <i>Pediatric:</i> ADI-R (87, 88), AQ-Adol-10 (82), AQ-Adol (89), AQ-C-10 (82), AQ-C (90), ASSQ (91), CARS2-HF (92, 93), CARS2-ST (92, 93), CAST (94), EQ-Adol (95), EQSQ (96), GARS-3 (97), M-Chat (98, 99), SCQ (100), SQ-Adol (95), SRS2 (101).
ADHD	9	418	<i>Adult:</i> ASRS-5 (102), ASRS-Checklist (103), CAARS (104), DIVA 2.0 (105), WURS (106). <i>Pediatric:</i> APA Inatt ¹ , Conners 3™ (107), DBDRS (108), NICHQ (109).
Schizophrenia	6	136	<i>Adult:</i> BPRS (110, 111), CGI-SCH (112), NSA-16 (113), PANSS (114), SANS (115–117), SAPS (118), BITE (119), EAT-26 (120), EDSS (121), EDE-Q (122), EDI-3 (123), SCOFF (124).

(Continued)

TABLE 1 | Continued

Disorder	Number of Questionnaires	Number of questions reviewed	List of Assessment Tools Included*
Cross-Category	16	6190	<i>Adult:</i> APA-CC-A ¹ , BSI (125), CIDI CAPI (126, 127), K10+ (128), MINI (129), PROMIS-Profile-A (130, 131), PROMIS-QB-A (130, 131), SCID-5-CV (3), SCL-90-R (132). <i>Pediatric:</i> APA-CC-C ¹ , CBC (133), DISC-IV (134), KSADS-PL-5 (135) PROMIS-Profile-C (130, 131), PROMIS-QB-C (130, 131), SDS (136).

*See **Supplementary Table 1** for further details and **Supplementary Table 5** for a list of abbreviations.

includes multiple APA anxiety assessments – see **Supplementary Table 1.

¹<https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures>.

together had a total of 10,154 questions. Where there were updated versions of a questionnaire, the latest or revised version of a questionnaire or interview was used although older versions were included if they were still commonly used. Assessment tools were obtained from the relevant publisher, publication or author and permission was sought where necessary. Of these assessment tools 82 were specifically designed for adults and 44 were specifically designed for pediatric populations. However, it is important to note that the adolescent age range falls within the adult range for some assessment tools, but within the pediatric range for others, depending on the specific questionnaire or interview. Where multiple versions of a questionnaire or interview were available, we selected the self-rated version. This was done to reflect the dominant prevalence of self-report tools in the mental health assessment literature, especially in relation to auxiliary tools, so as to strengthen comparisons between tools. Exceptions to this were for ASD and ADHD pediatric scales where the parent version was typically selected as this was the most commonly used approach for these disorders.

Table 1 provides a summary of the specific questionnaires included along with the total number of questions (see **Supplementary Table 5** for a list of abbreviations). The largest number of questionnaires or questions for a single disorder were for ASD followed by anxiety and depression.

Question Coding According to Symptom Category and Theme

Once we had selected these questionnaires and interviews, we then performed a systematic coding of each question within each questionnaire or interview to reflect the symptom(s) that each individual question assessed. This was done in two ways as described below.

Symptom Category Coding and Analysis

Firstly, the symptom(s) assessed in each individual question was identified and coded based on a judgement of the semantic content of the question. This resulted in 170 different preliminary symptom

codings. For example, if a question asked about a patient's difficulties with concentration or keeping focus on one thing, then the question was coded with the preliminary symptom coding of "difficulty concentrating", while if a question asked about difficulties falling asleep or about insomnia then it was coded with the preliminary symptom coding of "sleep problems".

These 170 preliminary symptom codings were then reviewed across all 10,154 questions and consolidated into a set of 43 master symptom categories which were used for the final analysis. This consolidation was performed by grouping together similar preliminary symptom codings, where those judged as being related were assigned to the same symptom category (for example, "concentration difficulties", "being easily distracted", "focusing on the big picture", and "mental fog" were all assigned to the category of "Attention, Concentration & Mental Focus"). The mapping of different preliminary codings to the symptom categories is provided in **Supplementary Table 2**.

The consolidation of the preliminary symptom codings into the 43 symptom categories required the grouping together of related symptoms into a single category. The choice to use a smaller number of symptom categories, rather than selecting all 170 preliminary symptom codings was both a practical one, and one which aimed to ensure our analysis didn't result in a high level of false positives where dissimilarities were over-emphasized. In this context, there may be some debate about the placement of a particular symptom within a common symptom category, based on an interpretation of their independence. While we consider the 43 symptom categories to be sufficiently distinct from one another in terms of the symptoms that they reflect, they may nonetheless arise from a common underlying cause in the same way that a fever and headache may arise from the same underlying etiology. Conversely, it is possible that similar symptoms grouped in the same symptom category may be interpreted in distinct ways by the respondent (for example "mental fog" and "difficulty concentrating") or arise from distinct causes. However, given that the underlying etiologies and relationships are not well established, mental health assessment, in general, arises at the level of psychological symptoms that are subject to semantic interpretation. By grouping together related symptoms, we believe we are more likely to have analyzed independent symptoms as being similar, rather than have analyzed similar symptoms as being dissimilar.

Within the dataset, 1.7% of questions (129) were marked as "excluded" because they either related to general functioning or to functions which were not specifically related to mental health and were not included in the analysis. For example, questions that asked about employment, hobbies or marital status were marked as "excluded", while questions which asked about general concerns, such as "What concerns you most about your child", were also excluded as they did not fit into any one symptom category. In addition, several of the cross-disorder tools were composed of multiple modules, with each module pertaining to a different disorder. In these instances, only modules which were related to the 10 clinical disorders of interest were included in the analysis and modules which specifically focused on other disorders were

excluded (see **Supplementary Table 1** for a list of included modules for the relevant interview assessments). This was done for practical reasons to focus the scope of the analysis towards the 10 disorders of interest. Where cross-disorder tools were not divided up into modules, all questions were included in the analysis. In addition, although the hierarchical nature of clinical questioning means that some questions within an interview are optional for the clinician (i.e., follow up questioning is dependent on the initial responses given by the patient), all interview questions were coded. Furthermore, while it may be the case that some questions within a questionnaire or interview do not contribute to a scoring algorithm or directly align with the symptom criteria which lead towards a formal diagnosis, they none the less would likely guide the clinician or researcher during the assessment and so all questions were included in the coding.

The coding of symptom category for each individual question was done in the following way: One researcher conducted an initial coding of all the questions by determining a preliminary symptom coding for each question that was then assigned to one of the 43 symptom categories as described above. A second researcher then reviewed, unblinded, these symptom category codings, across all 10,154 questions, and stated whether they agreed or disagreed with each coding. If they disagreed with a category coding for a particular question, they then generated an alternative coding. For each question where there was a disagreement in the coding between the two researchers, the difference was discussed between them, as well as reviewed by a third researcher to decide on the most appropriate coding for that question. This reassessment process was required for 0.3% of the questions.

For each questionnaire or interview, the proportion of questions corresponding to each of the 43 symptom categories was calculated. These counts were then converted into percentages to take into account the overall number of questions in that questionnaire or interview. This meant that the individual assessment tools could be more easily compared with one another without being biased by the fact that some assessment tools contained more questions than others. These percentages were then used in analysis of symptom similarity (see *Symptom Similarity Analysis*) and also averaged across the questionnaires and interviews corresponding to a particular disorder (e.g., all the ASD questionnaires and interviews) to provide aggregate views of the distribution of symptoms across each of the 10 disorders, as well as for the cross-disorder tools (**Figures 4B, C**).

Symptom Theme Coding and Analysis

The questions were then each coded according to the following "themes" representing the particular aspect of the symptom that was assessed: emotion (a feeling), cognitive (a thought), behavior (a behavior or action), physical (a physical or bodily symptom), trigger (the cause of a symptom), consequence (the outcome of a symptom), or treatment (medication or medical treatment relating to that symptom). The systematic coding of symptom themes was based on a semantic judgement of the content of the questions.

In this case both single and dual codings were used. Single coded items were used when questions asked only about one symptom

theme (e.g., how someone felt; whether they experienced a particular physical symptom). Dual codings were used to reflect the fact that some questions encompassed multiple themes such as a trigger and an emotion; a behavior and a thought. For example, a question that asked whether someone “tried hard not to think about it or went out of their way to avoid it” was dual coded as cognitive/behavior, while a question that stated “I have crying spells or feel like it” was dual coded as behavior/emotion. In addition, some symptoms more intuitively reflected dual themes. For example, “distressing flashbacks of unwanted memories” includes both emotion and cognitive elements and therefore were dual coded as emotion/cognitive. Similarly, “having a conversation” has both cognitive and behavioral elements and was dual coded as cognitive/behavior. This resulted in 30% of all questions being assigned a dual coding.

The coding of symptom themes was also carried out by two independent researchers in the same way as described above for symptom categories. For each question where there was a disagreement in the coding between the two researchers, the difference was reviewed between them, as well as by a third researcher to decide on the most appropriate coding for that question. This was necessary for 3.5% of the questions.

For each questionnaire or interview, the proportion of questions corresponding to each symptom theme was calculated. Single coded symptom themes were weighted as 1 and dual coded symptom themes were each weighted as 0.5. These counts were then converted into percentages to take into account the overall number of questions in that questionnaire or interview as was done for the symptom categories. These percentages were examined separately for each assessment tool and also averaged across the questionnaires and interviews within a particular disorder (e.g., all the ASD questionnaires and interviews) to provide an aggregate view of each of the 10 disorders, and for the cross-disorder tools (**Figure 3**).

Coding and Analysis of Other Question Features

In addition to coding the symptom category and symptom theme, other information embedded within the question content and/or answer options was also coded. This included coding whether the question asked about the presence, duration, severity, timing, or frequency of the symptom. For example, some questions focused on how often a patient experienced a particular symptom while other questions focused on the severity of the symptom experience. In the majority of cases, the answer options within the screening questionnaires provided an indication about the assessment style. For clinical interviews, where questions were often more open-ended, the wording of the question was used to infer the type of information that was being elicited. For example, questions starting with the words “how often” were considered to be assessing the frequency of a symptom, while questions asking about “how much” a symptom bothered someone was considered to be assessing symptom severity. For questions where the wording of the question was more ambiguous, or when a question simply asked whether a person was/had experiencing/experienced a particular symptom with no reference to timing, duration, frequency, or severity, the coding of “presence” was used.

Questions were also coded according to the time period over which the symptom was assessed. For example, some questions asked about symptoms occurring within the past week(s), past month(s), or past year(s) while others asked about symptoms occurring during childhood or across the entire lifetime. For the majority of screening questionnaires, this information was obtained from the introductory portion of the questionnaire. Where this information was not provided on the questionnaire or was more ambiguous, the questions were coded as “no specific time window” unless indicative information was specifically included in the individual questions. For the more open-ended clinical interviews, the time period information was determined based on information provided in the introductory sections of the interview modules, or from the individual question content itself. For example, a question may ask “In the last year...” was coded as “Past Year”, while “Have you ever...” and “did you ever have an episode when...” was coded as “Lifetime”, and “When you were 4/5 years old...” was coded as “Childhood”. Where the time period was more ambiguous in the interview question, or where no time period information was included, it was coded “no specific time window”.

Symptom Similarity Analysis

To determine the consistency of assessment within a disorder and the overlap between disorders, we computed a symptom similarity score between each pair of questionnaires both within and across disorders. Symptom similarity scores across pairs of assessment tools were calculated for each symptom category as the minimum percentage of questions dedicated to that category divided by the maximum percentage value. We then computed a weighted average using the mean of the two percentage values as the weighting factor to arrive at a questionnaire level similarity score. This meant that symptom level similarity scores were weighted more strongly when the symptom was associated with a greater proportion of questions overall within an assessment tool. We illustrate this with a simulated example with eight symptoms where there is a similar spread of symptoms across the pair of assessment tools (**Table 2**).

An overall similarity score was then computed for each disorder by averaging the similarity scores of each pair of assessment tools for that disorder. To compute the symptom overlap *between* disorders we averaged the scores between all pairs of assessment tools, where one of the pair was from each of the disorders (e.g., the average of tools designed to assess PTSD paired with tools designed to assess OCD).

RESULTS

Consistency of Assessment Formats

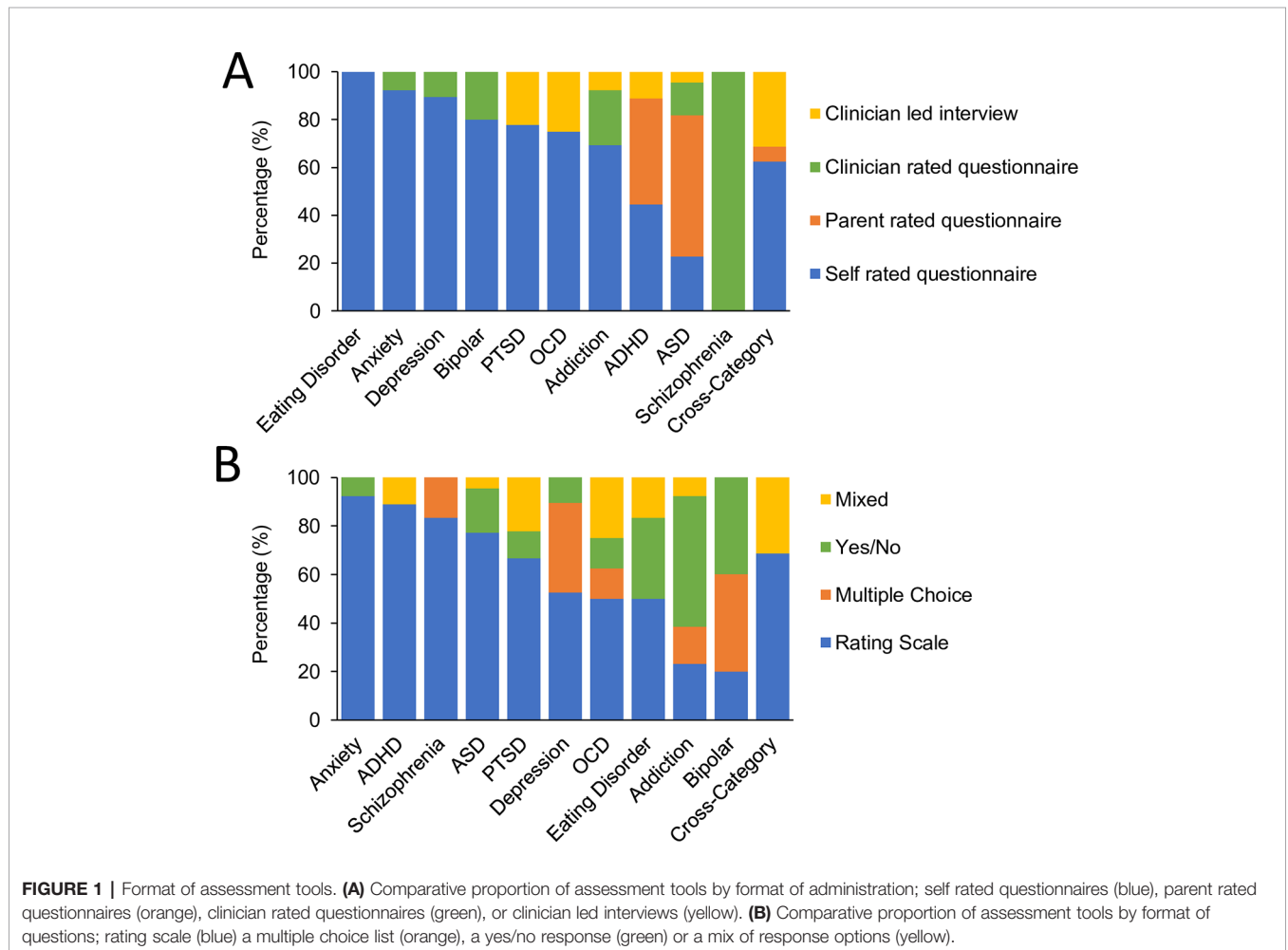
Out of the 126 questionnaires and clinical interviews selected for inclusion in the final analysis (**Table 1, Supplementary Table 1**) 80 (61 adult) were self-rated, 18 (all pediatric) were parent rated, 16 (14 adult) were clinician rated questionnaires and 12 (7 adult) were clinician led interviews. **Figure 1A** depicts the distribution of assessment types across the different disorders as well as for

TABLE 2 | Simulated example of symptom similarity computation.

Symptom	Questionnaire A	Questionnaire B	Symptom Similarity	Weighting Factor	Weighted Similarity
Symptom 1	12%	8%	66.7%	10%	6.67%
Symptom 2	12%	8%	66.7%	10%	6.67%
Symptom 3	12%	10%	83.3%	11%	9.17%
Symptom 4	12%	10%	83.3%	11%	9.17%
Symptom 5	12%	16%	75.0%	14%	10.50%
Symptom 6	12%	16%	75.0%	14%	10.50%
Symptom 7	14%	16%	87.5%	15%	13.13%
Symptom 8	14%	16%	87.5%	15%	13.13%
			Questionnaire Similarity		79%

the cross-category assessment tools. While most disorders were predominantly assessed by self-report, schizophrenia, ADHD, and ASD were predominantly assessed by the parent or clinician based on observation or opinion. This largely reflects the landscape of assessment. However, it also reflects our preferential selection of self-report tools over parent report where both were available for most disorders (e.g., 12 out of 19 self-report pediatric tools (63%) also had parent-rated versions which were not included) and the opposite for ADHD and ASD (1 out of 18 parent rated tools also had an alternative self-rating version which was not included).

In terms of question and answer format, overall 81 (48 adult) assessment tools were answered using rating scales, 20 (15 adult) were answered using a yes/no response option, 12 (11 adult) were answered using a multiple choice and 13 (8 adult) were answered using a mix of different answer formats corresponding to a more open ended style of interview. **Figure 1B** shows the distribution of answer formats across disorders as well as for the cross-category assessment tools. Rating scales dominated the assessment of most disorders while addiction stood out in being largely based on Yes/No answers thereby predominantly assessing presence rather than severity of symptoms. Cross



category, OCD, and PTSD were distinct in having a larger fraction of assessments using the mixed answer format associated with a more open-ended interview style.

Examination of other question characteristics revealed considerable variability in terms of whether questions assessed the presence, duration, severity, timing or frequency of the symptom. **Figure 2A** shows the percentage of questions within each assessment tool, averaged across each disorder, which asked about the presence, duration, severity, timing or frequency of the symptom. For schizophrenia, OCD, and PTSD, symptom severity were the predominant assessment formats. In contrast, the frequency of symptoms was the dominant assessment format for depression and anxiety. Assessments of symptom timing or duration were less common across all disorders.

When examining the time period over which symptoms were assessed, a similar pattern of heterogeneity was observed (**Figure 2B**). The most commonly assessed time period of symptoms for ADHD, PTSD, eating disorder, and OCD was the past few months while assessments of bipolar disorder, anxiety and depression more commonly used a time period of the past few days or weeks. In contrast, assessments of ASD did not readily specify a

time period for symptoms while addiction assessments either did not specify a time period or used a time period covering the past year(s). There was therefore very little consistency in assessment time period across the different disorders.

Overview of Symptom Categories, Themes, and Characteristics

Each assessment tool was analyzed along various dimensions to reveal the distribution of symptom themes and symptom categories. **Figure 3** shows the proportion of questions corresponding to each symptom theme averaged across assessments for each disorder. The results showed that in the aggregate, depression alone was dominated by symptoms that related to emotion (57% of questions). The remaining questions were distributed across physical and cognitive symptoms (20% and 15% respectively). In contrast ADHD was dominated by questions relating to behavioral symptoms (51%) while only 13% focused on emotion related symptoms. Others such as OCD, ASD, eating disorder, and schizophrenia were distributed more uniformly across multiple themes. Assessments for OCD, PTSD and anxiety more commonly asked about the triggers associated with a symptom, while

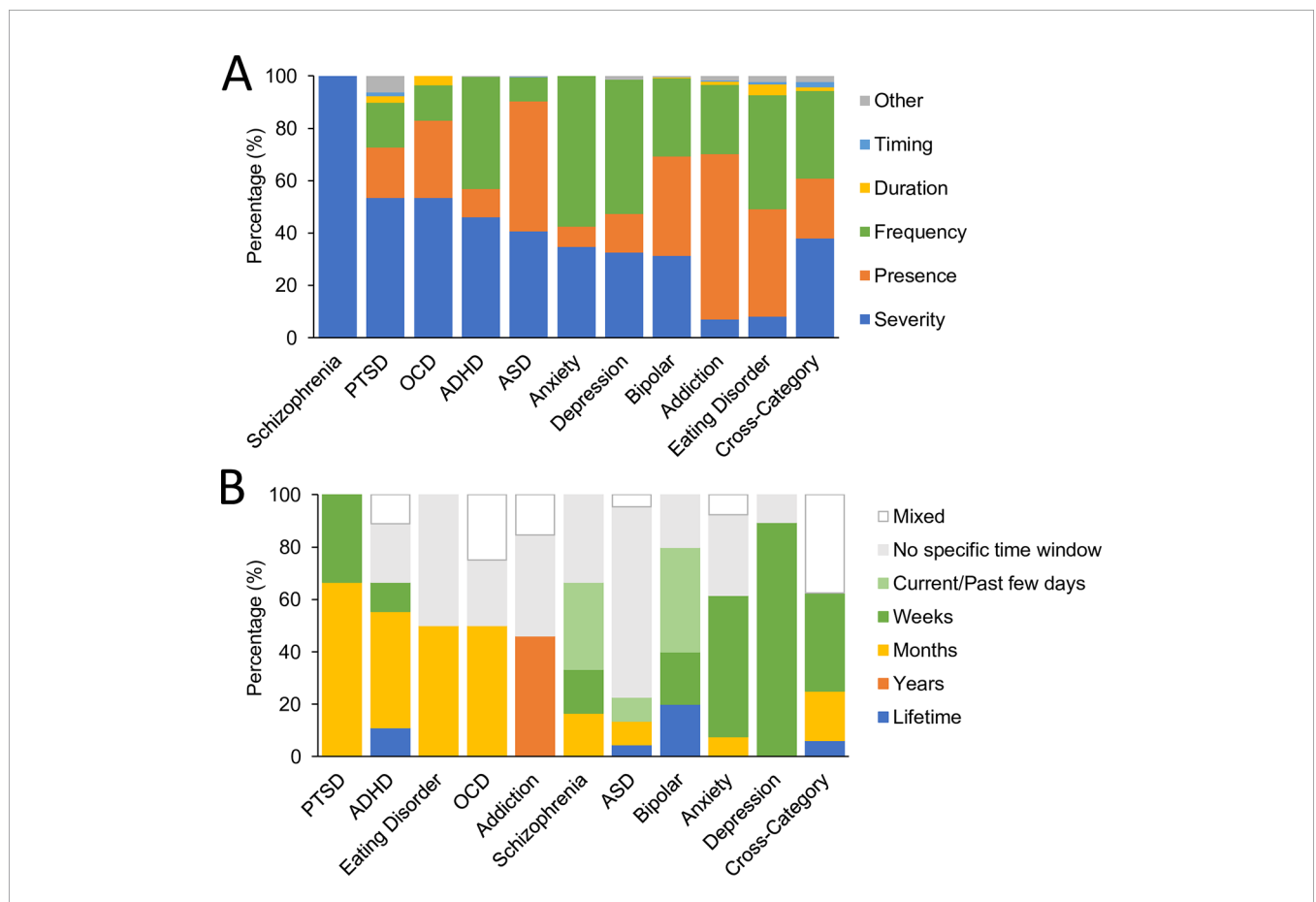
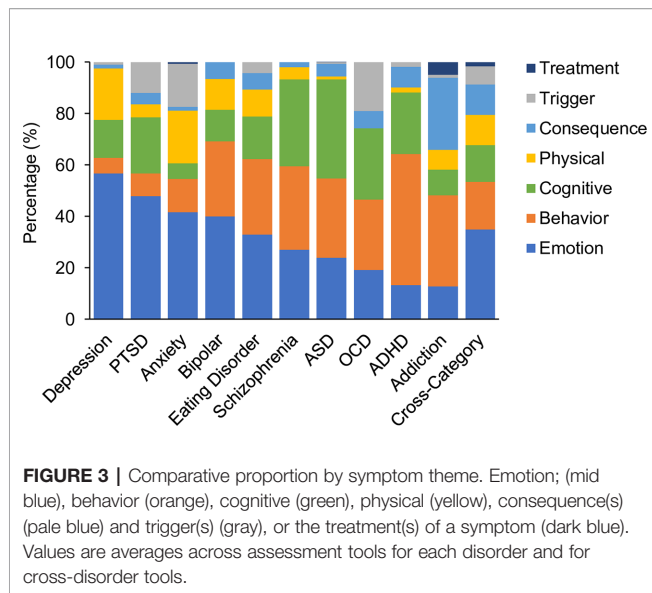


FIGURE 2 | Symptom assessment characteristics. **(A)** Comparative proportion of questions within each disorder and for cross-disorder tools (averaged across assessment tools) by symptom aspects; presence (orange), duration (yellow), severity (dark blue), timing (light blue), or frequency (green) of the symptom. **(B)** Comparative proportion of assessment tools by time period of symptom assessment; currently/past few days (includes past 24 h and past 48 h; pale green), weeks (includes past week and past 2 weeks; dark green), months (includes past month, past 3 months and past 6 months; yellow), years (includes past year and past 2 years; orange), lifetime (includes childhood, adulthood, lifetime and lifetime episodes, blue), no specific time window (gray), and mixed (white).



assessments of depression, bipolar, ASD, schizophrenia, addiction, and ADHD rarely asked questions of this type. Assessments of addiction and ADHD more commonly asked about the consequences of a symptom, but this was more rarely assessed for other disorders. Symptom treatments were most commonly asked about during assessments of addiction. As might be expected, cross-disorder assessments included questions which covered all symptom themes, with the greatest proportion of questions asking about emotions.

Figure 4A shows the percentage representation of each of the 43 symptom categories across each of the disorders as well as for the cross-disorder questionnaires (also see **Supplementary Table 3**). The primary takeaway is the large degree of symptom overlap across disorders. However, a number of symptom categories were dominant in only one disorder. For example, “Substance use and Addiction” was dominant for addiction (65% of addiction assessment) as well as being a symptom assessed in bipolar disorder. In contrast, other symptom categories such as “Fear, Panic & Anxiety”, “Mood & Outlook”, “Confidence & Self-Judgement”, “Interpersonal”, and “Attention & Concentration & Mental Focus” dominated assessments of multiple disorders. Fear, Panic, & Anxiety for instance was highest for anxiety (37%) but was also substantially represented in the related disorders of PTSD (15%) and OCD (17%). The percentage of disorders represented by different symptom categories are shown in **Figure 4B** (excluding cross-category assessment tools). Overall three symptom categories (7%) were represented across all disorders and 4 (9%) were represented in all but one disorder. In the aggregate, 42% of symptom categories were represented in over half of all disorders. There was only one symptom category that was assessed by only one disorder (Allergies & Rashes in ADHD assessment), representing a physical symptom that may or may not be specifically relevant to a mental health disorder.

Conversely, all disorders encompassed multiple symptom categories. The distribution is shown in **Figure 4C**. OCD was the most symptom specific covering only 39% of symptom categories

followed by eating disorder (42%), addiction and PTSD (both 47%). In contrast depression was the most general, encompassing 70% of all symptom categories followed by ASD and ADHD (both 67%). While the weighting of these overlapping symptoms certainly differs across disorders, it nonetheless demonstrates that no individual symptom category can be considered a definitive element of any disorder categorization.

Similarity Scores Within and Across Disorders

The previous sections illustrate the large heterogeneity in assessment format as well as the substantial overlap in the range of symptoms across disorders. However, the particular weighting of symptoms in the assessments varied (i.e., the number of questions dedicated to a particular symptom category). Therefore, to get a better sense of the distinctions between disorders along this dimension, we computed a measure of symptom consistency across disorders (see Materials and Methods). **Figure 5** shows a (symmetric) matrix of the average similarity scores within and between disorders (values pertaining to this figure can be found in **Supplementary Table 3**).

Analysis of similarity scores between pairs of disorders ranged from 2% to 28%. The most similar assessments (i.e., above 20%) from a cross-disorder perspective were between PTSD and OCD, depression and PTSD, anxiety and PTSD, PTSD and eating disorder, ASD and schizophrenia, depression and bipolar disorder, and bipolar disorder and schizophrenia. The disorder pairs that had the lowest similarity scores (under 4%), and therefore were most dissimilar in terms of symptoms considered, were OCD and ASD, OCD and bipolar disorder, OCD and schizophrenia, ASD and addiction, and schizophrenia and addiction. Thus, while the range of symptoms overlapped considerably across many disorders, the distinctions between them were largely at the level of importance ascribed to the symptom.

While on the face of it, this appears to support the specific divisions of disorders, the disorder comparison analysis was done based on an average of the tools within any individual disorder. However, when looking at the overall consistency of assessment of any individual “disorder”, symptom similarity was surprisingly low (diagonal of the matrix in **Figure 5**). OCD showed the greatest level of assessment consistency (58% similarity across 8 tools), followed by addiction (51% across 13 tools). The other disorders showed much lower consistency with assessments of bipolar disorder and schizophrenia being the lowest at 29%. Thus, when there is substantial symptom ambiguity within disorders, it calls into question the relevance of comparing between two disorders (as we have done above) as individual assessment tools can have vastly diverging similarity scores. For example, the overall similarity between depression and ADHD was 16%, but when looking at individual pairs, PHQ-9 (depression), and CAARS (ADHD) were 37% similar, while PHQ-9 (depression) and ASRS-5 (ADHD) were only 8% similar.

Individual Disorders

We next present the consistency across individual assessment tools within each disorder separately. The consistency of symptom themes and categories among individual assessments, as well as the similarity scores, are reported. For completeness,

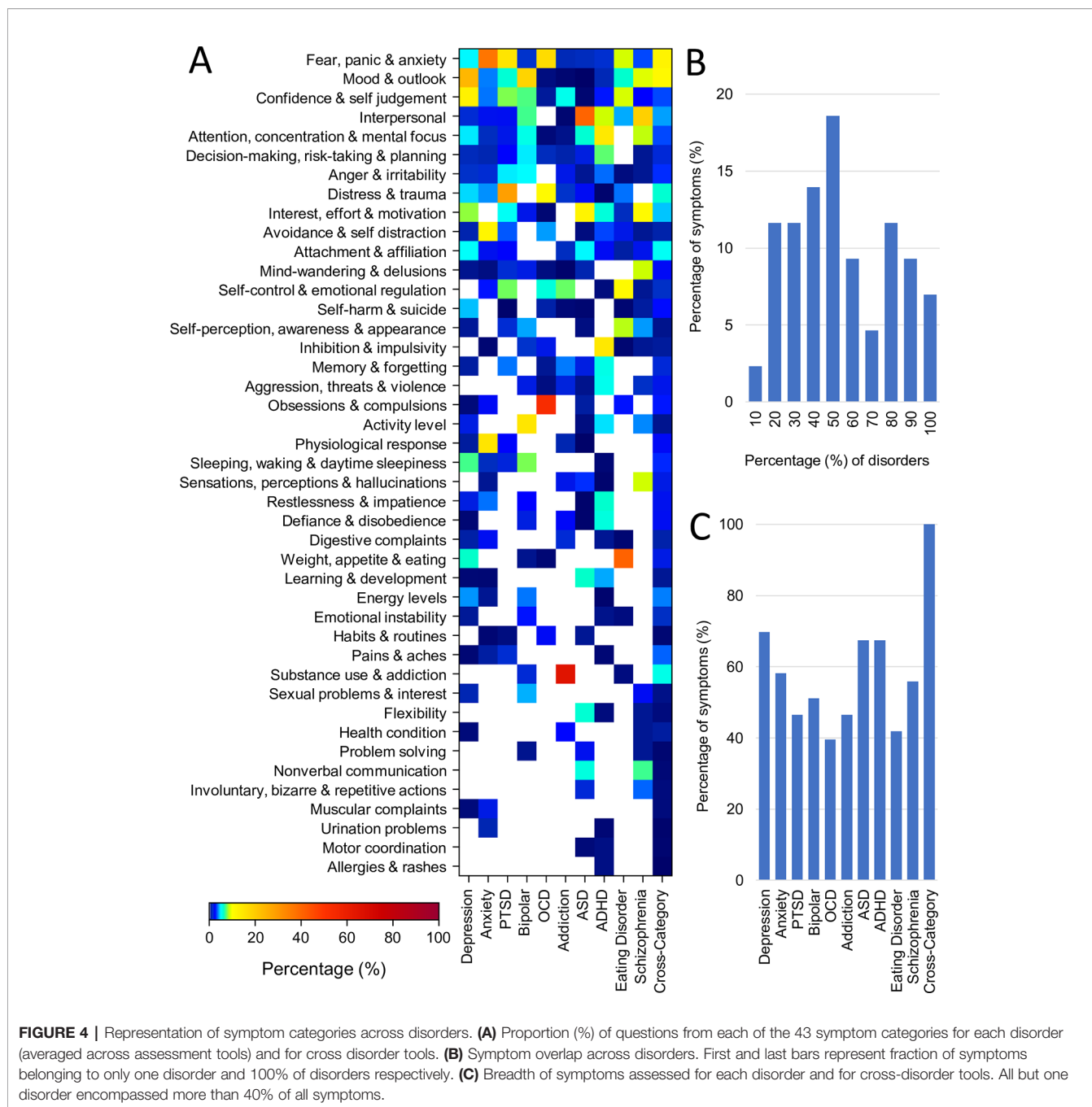


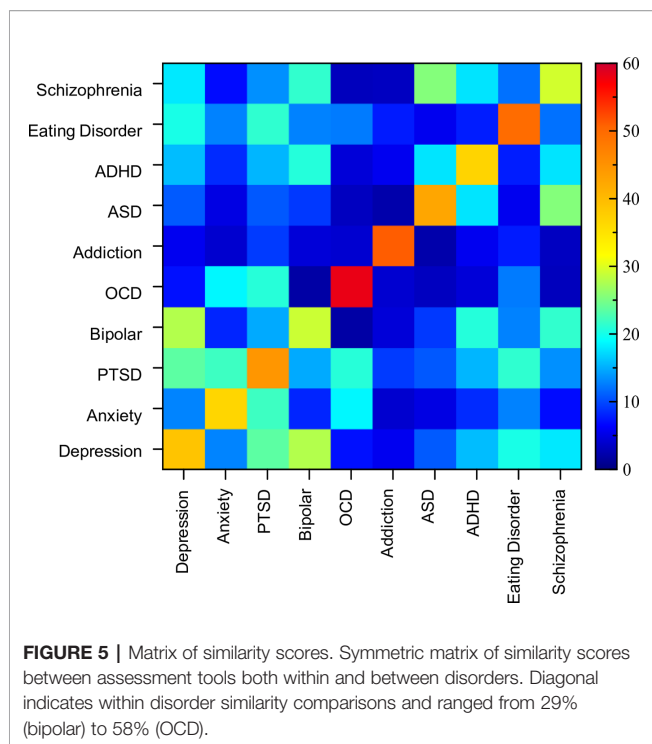
FIGURE 4 | Representation of symptom categories across disorders. **(A)** Proportion (%) of questions from each of the 43 symptom categories for each disorder (averaged across assessment tools) and for cross disorder tools. **(B)** Symptom overlap across disorders. First and last bars represent fraction of symptoms belonging to only one disorder and 100% of disorders respectively. **(C)** Breadth of symptoms assessed for each disorder and for cross-disorder tools. All but one disorder encompassed more than 40% of all symptoms.

we also report the breadth of symptoms, with the caveat that this metric is influenced by the total number of questions within an assessment tool, especially when the number of questions were lower overall. For example, an assessment tool with only 15 questions cannot cover a wide breadth of symptoms, while an assessment tool with 700 questions could potentially do so.

Depression

Nineteen auxiliary depression assessment tools were analyzed. Thirteen of these were predominantly developed for adult populations and 6 for pediatric populations (see **Table 1**,

Supplementary Table 1). Two other common assessments of depression (the DASS: Depression, Anxiety Stress Scales (137, 138); and HADS: Hospital Anxiety and Depression Scale; (139) were excluded from this study as they spanned two disorders of interest. **Figure 6A** shows the considerable variability of depression assessments in terms of symptom themes (see also **Supplementary Table 3**). For example, questionnaires such as HAMD, QIDS, and IDS included a lower proportion of emotion themed questions (ranging from 21% to 37%) and a higher proportion of physical themed questions (ranging from 43% to 56%), compared to the other assessment tools. In contrast, the APA-Dep-C, APA-Dep-A,



RADS2, CESDC, CES-D, and GDS-SF had a greater proportion of emotion themed questions (ranging from 70% to 100%). None of the questionnaires prioritized cognitive or behavioral symptoms within their questions (a maximum of 29% for cognitive symptoms and a maximum of 19% for behavioral symptoms across all depression assessments). In addition, none of the depression assessments asked about previous or current treatments, and only 37% and 26% of assessments asked about symptom triggers and consequences respectively.

Symptom categories also varied considerably across assessments. Seventy percent of all symptom categories were represented across all the assessment tools together. However, only 10% of these were common across 80% or more assessment tools (Confidence & Self judgement; Mood & Outlook; Interest, Effort & Motivation). The similarity analysis showed that the depression assessments were, on average, only 39% similar. One of the tools (HAMD) included 4 questions which were not used in the scoring for depression. When HAMD was removed from the analysis, the average similarity score moved up by one percentage point to 40%. Similarity scores between individual pairs of assessments ranged from 7% (APA-Dep-A and QIDS) to 94% (CES-D and CES-DC). Comparisons between adult and pediatric tools showed that adult tools were, on average, 40% similar, pediatric tools were 46% similar, and the similarity score between adult and pediatric tools was 36%, suggesting there were some differences between the two groups. See **Supplementary Table 4** for a table of similarity scores between each depression assessment tool.

Anxiety Assessments

Thirteen auxiliary anxiety assessment tools were included in the analysis of which 8 were for adult populations and 5 were pediatric

(see **Table 1** and **Supplementary Table 1** for details). One additional common anxiety assessment (RCADS, Revised Children's Anxiety and Depression Scale) was excluded from this study due to its cross-disorder approach. We note that two assessment tools (APA-Anx-A and APA-Anx-C) were composed of a set of related APA assessments each focusing on different types of anxiety (covering generalized anxiety, panic, agoraphobia, social phobia, separation anxiety and social anxiety which we have aggregated together as one broad assessment of anxiety). In addition, 2 assessment tools (LSAS and SPIN) included a greater number of questions on social anxiety. We performed the analysis both with and without these two assessment tools to determine whether including them skewed the results in any way but found only one minor change in the results (see below) so we included them in the analysis.

Looking across all the anxiety tools, emotion themed symptoms dominated in GAD7 and STAI (79% and 85% respectively) while BAI, HAMA, and ZAS emphasized physical symptoms (ranging from 55% to 67% of questions; **Figure 6B**, **Supplementary Table 3**). Many of the assessment tools (62%) also asked about symptom triggers. In general, the pediatric tools were more comprehensive in terms of covering all symptom themes (with the exception of treatment) compared to the adult tools. For example, all but one of the pediatric tools (but only 1 of the adult tools) asked about consequences of a patient's symptoms. In addition, all pediatric tools included questions about cognitive and behavioural symptoms, while there were gaps and heterogeneity across the adult assessment tools.

Fifty-eight percent of symptoms were covered across all anxiety assessment tools. However, only 1 symptom (Fear, Panic, & Anxiety) was assessed by 80% or more of the assessment tools. This resulted in, on average, only a 37% similarity across all assessment tools. One reason for this relatively low similarity score could be due to the fact that anxiety has multiple dimensions (e.g. phobia, separation anxiety, social anxiety) which could be differentially assessed across tools. In addition, some assessments asked about specific symptoms occurring over the past few weeks (GAD7), whilst others (especially for pediatric tools) were more trait-like and asked about the degree to which an individual generally experienced symptoms associated with anxiety (e.g., RCMAS), which could also have impacted the overall similarity score. Similarity scores between individual pairs of assessments ranged from 6% (LSAS and ZAS) to 86% (APA-Anx-C and APA-Anx-A; See **Supplementary Table 4** for details). One of the tools (SCAS) included 6 questions which were not used in the scoring for anxiety. However, when SCAS was removed from the analysis, the average similarity score did not change. To explore the impact of including LSAS and SPIN we examined the average similarity scores between these two assessment tools and all the other assessment tools and showed that they were, on average, 36% similar to the other tools, which was not very different from the average overall similarity score of 37% above. Comparisons between adult and pediatric tools showed that adult tools were, on average, 31% similar, pediatric tools were 53% similar, and the similarity score between adult and pediatric tools was 37%, suggesting that pediatric tools were considerably more similar than adult tools.

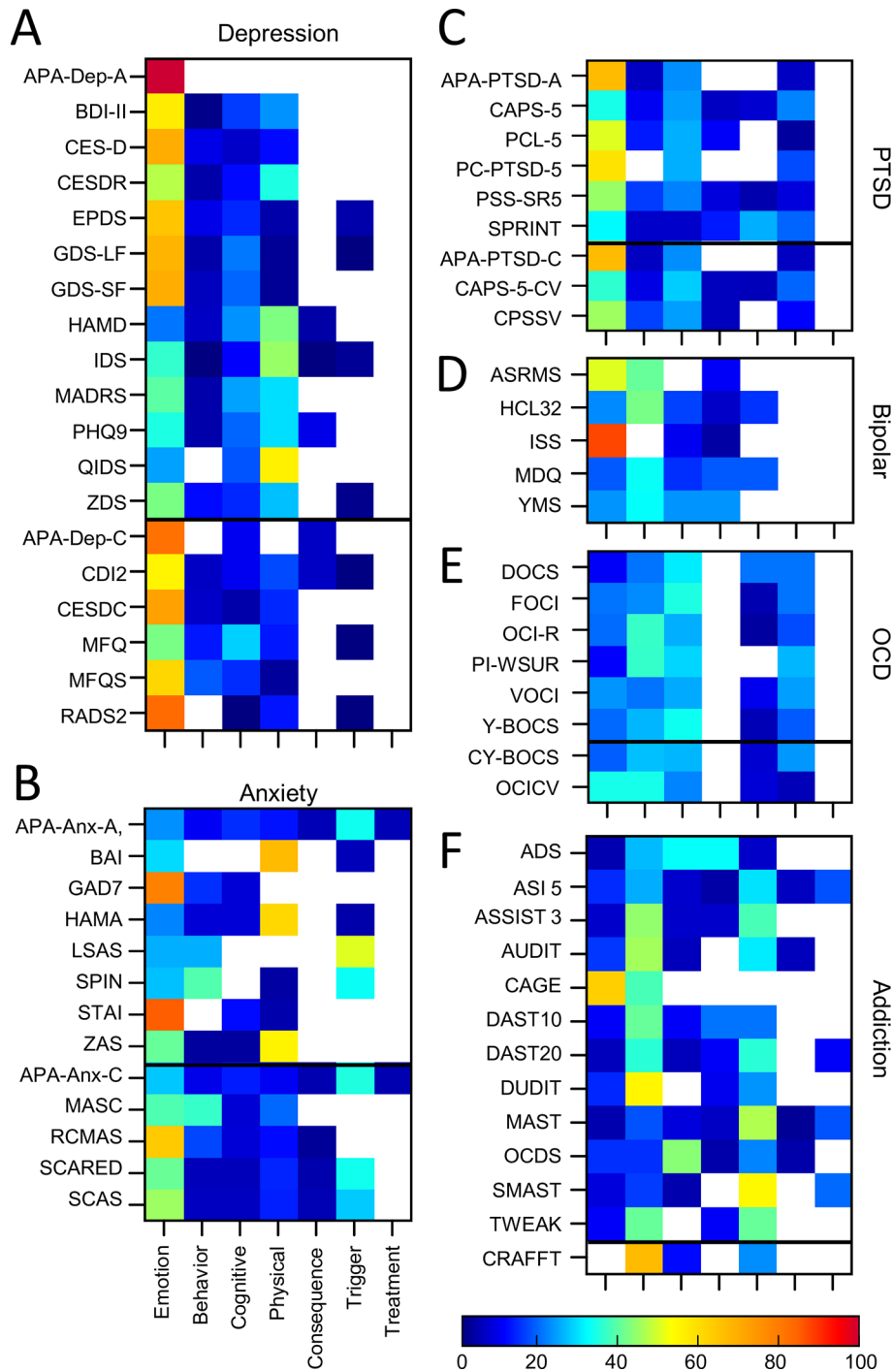


FIGURE 6 | Proportion of symptom themes. Proportion of symptom themes (%) across individual assessment tools within individual disorders. **(A)** Depression. **(B)** Anxiety. **(C)** PTSD. **(D)** Bipolar. **(E)** OCD. **(F)** Addiction. Black line denotes delineation between adult (top) and pediatric (bottom) tools. See **Supplementary Table 5** for a list of abbreviations.

PTSD Assessments

Nine PTSD assessment tools were included; six assessments for adult populations, and three assessments for pediatric populations (see **Supplementary Table 1** for details). Two

tools were diagnostic (clinician led) interviews, whilst the remaining were auxiliary (self report). Emotion themed symptoms dominated most PTSD assessment tools (ranging from 31% to 67%; **Figure 6C**, **Supplementary Table 3**).

Furthermore, with the exception of PC-PTSD-5, all assessment tools also included cognitive, behavioral and trigger themed symptoms. Six out of the nine assessment tools also included physical themed symptoms and four assessments considered consequences. None asked about symptom treatment.

Forty-seven percent of symptoms were asked about in total across all PTSD assessments of which only 15% were common across 80% or more assessment tools (Confidence & Self judgement; Fear, Panic & Anxiety; Distress & Trauma). This resulted in a 45% similarity score on average across all PTSD assessments. Similarity scores between individual pairs of assessments ranged from 16% (PCL-5 and SPRINT) to 100% (APA-PTSD-A and APA-PTSD-C; See **Supplementary Table 4** for details). We also examined the impact of including both diagnostic and auxiliary tools on the similarity scores and showed that the diagnostic interviews were 45% similar to the auxiliary tools, while the similarity scores of the auxiliary tools alone was 43%. Comparisons between adult and pediatric tools showed that adult tools were, on average, 40% similar, the 3 pediatric tools were 44% similar, and adult and pediatric tools were 48% similar, influenced by the fact that two pediatric tools had corresponding adult versions.

Bipolar/Mania Assessments

Five bipolar/mania assessment tools were included in the analysis, all adult (See **Supplementary Table 1** for details). Of these, ISS and ASRMS covered a greater proportion of emotion themed symptoms compared to the other assessment tools (88% and 50% respectively; **Figure 6D**, **Supplementary Table 3**). The scales varied in terms of the proportion of questions that asked about behavioural symptoms (ranged from 0% to 41%), cognitive symptoms (ranged from 0% to 23%) and physical symptoms (ranged from 3% to 23%). None of the assessment tools asked about symptom treatments or triggers, while two out of the five assessment tools asked about consequences of the symptoms.

Fifty-one percent of symptoms were asked about in total, of which only 23% of symptoms were assessed by 80% or more of the assessment tools (Activity Level; Sleeping, Waking & Daytime Sleepiness; Confidence & Self judgement; Mood & Outlook; Anger). On average, bipolar assessments were 29% similar, among the lowest consistency of any disorder covered here. Here again, one tool (HCL32), used only 32 questions (out of a possible 43) in the actual scoring. When HCL32 was removed from the analysis then the average similarity score decreased to 26%. Similarity scores between individual pairs of assessments ranged from 16% (ISS and YMS) to 41% (ISS and HCL32; see **Supplementary Table 4** for details).

OCD Assessments

Eight OCD assessments were analyzed. This included 6 for adult populations and two for pediatric populations (See **Supplementary Table 1** for details). **Figure 6E** shows the distribution of symptom themes across each OCD assessment tool showing a roughly even proportion of questions across multiple themes and a general consistency in their proportions across the assessments (see also **Supplementary Table 3**). All the OCD assessment tools contained questions themed as emotion

(ranged from 10% to 33%), cognitive (ranging from 21% to 34%) and behavioral (ranged from 20% to 36%) symptoms. All assessment tools also asked about symptom triggers (ranged from 5% to 26% of questions) and all but one asked about consequences (ranged from 3% to 20% of questions). None of the assessment tools included in this analysis asked about physical symptoms or treatment.

The symptom categories across all OCD assessment tools were relatively compact encompassing only 40% of all symptoms. 12% of these were found in 80% or more of the assessment tools (Obsessions & Compulsions; Distress & Trauma). OCD assessment also had the highest similarity on average of all disorders (58%). For both YBOCS™ and CYBOCS™, only the scale element (excluding 2 questions) and not the checklist element of the tool contributed to the actual scoring algorithm. When these assessments were excluded from the analysis, the average similarity score decreased to 55% and the two clinician led assessment tools were, on average, 59% similar to the other PTSD assessment tools. Similarity scores between individual pairs of assessments ranged from 41% (PI-WSUR and OCI-R) to 90% (Y-BOCS™ and CY-BOCS™; see **Supplementary Table 4** for details). Comparisons between adult and pediatric groups showed that adult tools were, on average, 55% similar, the two pediatric tools were 54% similar, and adult and pediatric tools were 63% similar, influenced by the fact that the two pediatric tools had corresponding adult versions.

Addiction Assessments

Thirteen addiction assessment tools were included of which 12 were developed primarily for adult populations and one for pediatric (adolescent) populations (see **Supplementary Table 1** for details). Seven of these focused on alcohol addiction (ADS, AUDIT, CAGE, MAST, OCDS, SMAST, TWEAK), 3 focused on drug addiction (DUDIT, DAST10, DAST20), and 3 focused on both alcohol and drug addiction (CRAFT, ASI5, ASSIST3). Assessments pertaining to other forms of addiction (e.g., gambling, gaming) were not included so as to limit the scope to the most common forms of clinical addiction.

There was considerable heterogeneity in symptom themes when looking across all addiction assessments (**Figure 6F**, **Supplementary Table 3**). For example, two tools included questions which covered all seven themes (the MAST and ASI-5) while other assessment tools only covered two or three themes (e.g., CAGE, CRAFT). The majority of the addiction assessments included here focused primarily on behavioral symptoms (ranging from 14% of questions for OCDS through to 67% of questions for CRAFT). However, a number of assessment tools also focused strongly on emotional symptoms (e.g., CAGE, 63% of questions), cognitive symptoms (e.g., OCDS, 43% of questions) or physical symptoms (e.g., ADS, 32% of questions). All but one tool asked questions on symptom consequences, and a number also included questions which asked about symptom triggers and treatment.

When examining the breadth of symptoms assessed across all addiction assessment tools, the results show that 47% of all symptoms were covered across all assessments of which only 1 symptom was common across 80% or more assessment tools

(Substance Use & Addiction). However, despite this, on average the addiction assessment tools included here were 51% similar. Similarity scores between individual pairs of assessments ranged from 19% (TWEAK and ASI-5) to 89% (ASSIST3 and SMAST; See **Supplementary Table 4** for details). When assessing alcohol addiction and drug addiction separately, the average similarity scores was 51% between alcohol addiction assessments, 57% between drug addiction assessments and 38% between tools assessing both drug and alcohol addiction. The ASI clinical interview which assessed both drug and alcohol addiction included a section at the end on general psychiatric status and showed the greatest dissimilarity from the other addiction assessments (it was on average only 25% similar to all the other addiction assessment tools). When the ASI was excluded from the analysis the overall average similarity score went up from 51% to 55%.

ASD Assessments

Twenty-two ASD assessment tools were included of which 6 were primarily developed for adult populations and 16 assessment tools for pediatric populations (see **Supplementary Table 1** for details). All assessments included emotion, cognitive and behavioural themed symptoms (**Figure 7A, Supplementary Table 3**). However, there was considerable heterogeneity in terms of which theme was dominant in any one assessment tool. For example, approximately half of the assessment tools comprised of questions where the assessment of behavioral symptoms was dominant. In contrast, the other ~half of assessment tools comprised of questions where the assessment of cognitive symptoms was dominant (and this included all adult tools). Across all ASD assessment tools, the proportion of questionnaires that focused on cognitive symptoms ranged from 5% (M-CHAT) to 80% (AQ-A-10), while the proportion of questionnaires which focused on behavioral symptoms ranged from 5% (EQ-A) to 71% (CARS2-ST) indicating the high level of variability across the different assessment tools. In addition, only one tool, the EQSQ, predominantly focused on emotional symptoms (42%).

Sixty-seven percent of symptoms were evaluated in total across all ASD assessments. Of these, only 10% of symptoms were common across 80% or more assessment tools (Attachment & Affiliation; Interpersonal; Interest, Effort & Motivation). This resulted in a similarity score of only 43% overall, demonstrating a fairly high level of inconsistency. One questionnaire (CAST) included seven questions (out of 38) which were not used for the scoring. When CAST was excluded from the analysis, the average similarity score shifted by one percentage point to 42%. Similarity scores between individual pairs of assessments ranged from 3% (M-CHAT and SQ-A) to 100% (AQ-Adol and AQ-C; AQ-Adol and AQ-A; AQ-A and AQ-C; EQ-A and EQ-Adol; see **Supplementary Table 4** for details). The average similarity between the diagnostic assessment tool (ADI-R) and all the other auxiliary tools was 41%, similar to the overall similarity score of 43%. In addition, as the adult ASD tools were predominantly self-rated and the pediatric tools were predominantly parent-rated we compared these two groups separately. The average similarity scores were as follows: 39%

within the adult tools; 44% within the pediatric tools; and 42% between the adult and pediatric tools.

ADHD Assessments

Nine ADHD assessment tools (eight auxiliary) were included of which five were for adult populations and four were for pediatric populations (see **Supplementary Table 1** for details). Three of these pediatric tools (Conners3™, DBDRS, and NICHQ) were also designed to assess symptoms relating to conduct disorder and oppositional defiance disorder.

The majority of the ADHD assessment tools focused on behavioral symptoms (ranging from 28% to 81% of questions; **Figure 7B, Supplementary Table 3**). This was most apparent for DBDRS, ASRS-5, and NICHQ (with 81%, 58%, and 57% of questions, respectively). Questions assessing cognitive symptoms were also common in the majority of assessment tools (ranging from 9% for NICHQ through to 44% APA-Inatt). The inclusion of questions which assessed emotional symptoms was relatively low (ranging from 0% for APA-Inatt to 24% for CAARS) and only three out of the nine tools asked about physical symptoms. All tools but two (ASRS-5, CAARS) asked about both triggers and consequences of the symptoms although none asked about symptom treatment.

Across all ADHD assessment tools 67% of symptoms were asked about in total of which 21% of symptoms were assessed by 80% or more of the assessment tools (Attention & Concentration & Mental Focus; Decision-making, Risk taking & Planning; Restlessness & Impatience; Inhibition & Impulsivity; Interpersonal; Interest, Effort & Motivation). However, there was only 37% similarity overall across ADHD assessment tools. Similarity scores between individual pairs of assessments ranged from 11% (ASRS-Checklist and WURS) to 68% (ASRS-Checklist and DIVA 2; See **Supplementary Table 4** for details). In addition, one questionnaire (WURS) used only 25 (out of 61) questions in the scoring algorithm for ADHD while another (ASRS-Checklist) only used the first 6 questions (out of 18) for scoring. Exclusion of the WURS from the analysis increased the overall average similarity score to 40%, while exclusion of the ASRS-Checklist resulted in no change. The average similarity score across the auxiliary tools only (i.e., without DIVA 2) was 35%. Furthermore, to determine the impact of including assessment tools that considered symptoms for conduct disorder (CD) and oppositional defiance disorder (ODD), we examined the similarity scores for these two groups separately. The average similarity score between the +CD/ODD and -CD/ODD groups of assessment tools was 32%, while the average similarity within the -CD/ODD group (6 assessment tools) was 38% and the average similarity within the +CD/ODD group (three assessment tools) was 63%. In addition, as the adult ADHD tools were predominantly self-rated (except DIVA-2) and the pediatric tools were predominantly parent-rated we compared these two groups separately. The average similarity scores were 40% within the adult tools, 42% within the pediatric tools and 34% between the adult and pediatric tools indicating that there may be some differences between adult and pediatric tools.

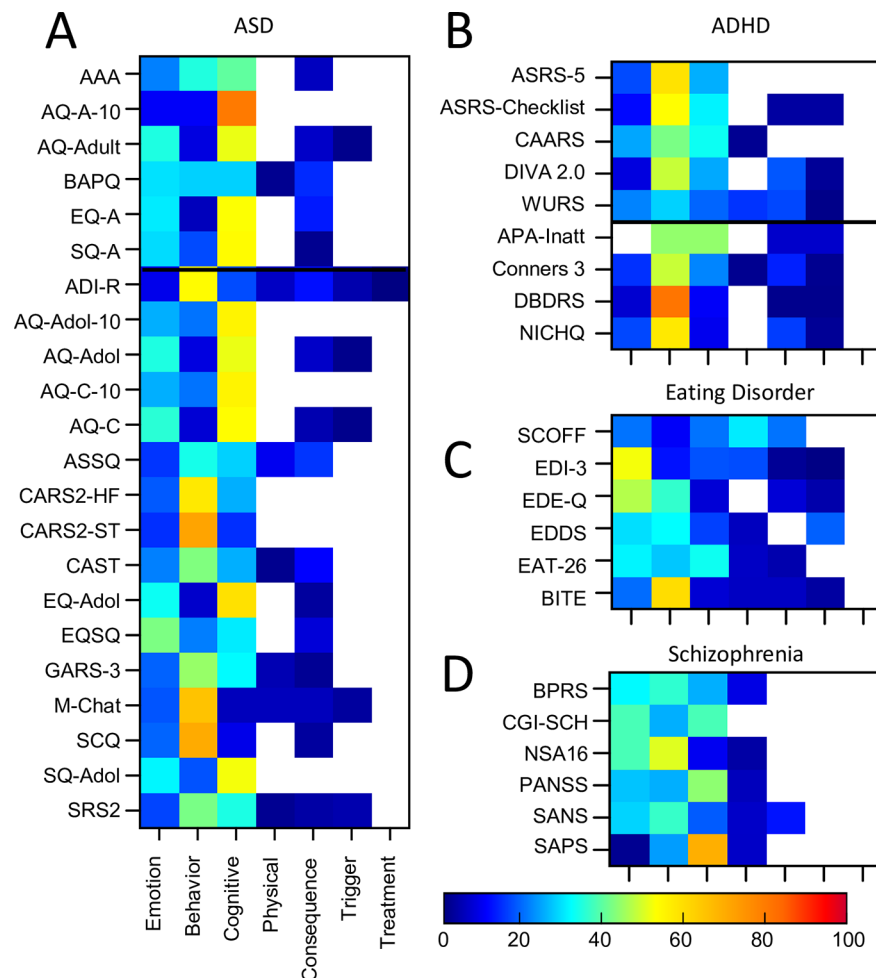


FIGURE 7 | Proportion of symptom themes. Proportion of symptom themes (%) across individual assessment tools within individual disorders (continued). **(A)** ASD. **(B)** ADHD. **(C)** Eating Disorder. **(D)** Schizophrenia. Black line denotes delineation between adult (top) and pediatric (bottom) tools. See **Supplementary Table 5** for a list of abbreviations.

Eating Disorder Assessments

Six eating disorder assessment tools were included (all adult/adolescent; all auxiliary; see **Supplementary Table 1** for details). The majority of the tools focused on behavioral (ranging from 10% to 60% of questions) and emotional symptoms (ranging from 19% to 52% of questions; **Figure 7C, Supplementary Table 3**). A number of tools (EAT26, EDI-3, EDDS) also focused on cognitive symptoms (33%, 17%, 16% respectively). With the exception of EDI (17%) and SCOFF (30%), the inclusion of questions which assessed physical symptoms was typically lower. All but 2 tools (SCOFF and EAT26) asked about triggers and all but 1 (EDDS) asked about consequences while no tools asked about symptom treatment.

Across all eating disorder assessment tools 42% of symptoms were asked about in total of which 17% of symptoms were assessed by 80% or more of the assessment tools (Fear, Panic & Anxiety; Confidence & Self Judgement; Self-perception,

Awareness & Appearance; Weight, Appetite, & Eating; Self Control & Emotional Regulation). There was 50% similarity overall across eating disorder assessment tools and similarity scores between individual pairs of assessments ranged from 30% (EAT-26 and SCOFF) to 79% (BITE and EDDS; See **Supplementary Table 4** for details).

Schizophrenia Assessments

Six clinician rated schizophrenia assessment tools for adult populations were included (all auxiliary; See **Supplementary Table 1** for details). **Figure 7D** shows there was considerable variation across the themes dominating these tools as follows: emotional (1% to 38%), cognitive (9% to 69%), behavioral (24% to 50%), and physical (0% to 8%; see also **Supplementary Table 3**). There was rarely a focus on the consequences, triggers, or treatment of the symptoms.

Fifty-six percent of symptoms were covered across all assessments of which only 13% of symptoms were common to 80% or more of the assessment tools (Attention & Concentration & Mental Focus; Interpersonal; Self-perception, Awareness & Appearance). The 6 assessment tools were, on average, only 29% similar. The differential focus on positive vs negative symptoms of schizophrenia across assessments in part contributed to the dissimilarity across assessments. Similarity scores between individual pairs of assessments ranged from 8% (CGI-SCH and PANSS) to 52% (BPRS and PANSS; See **Supplementary Table 4** for details).

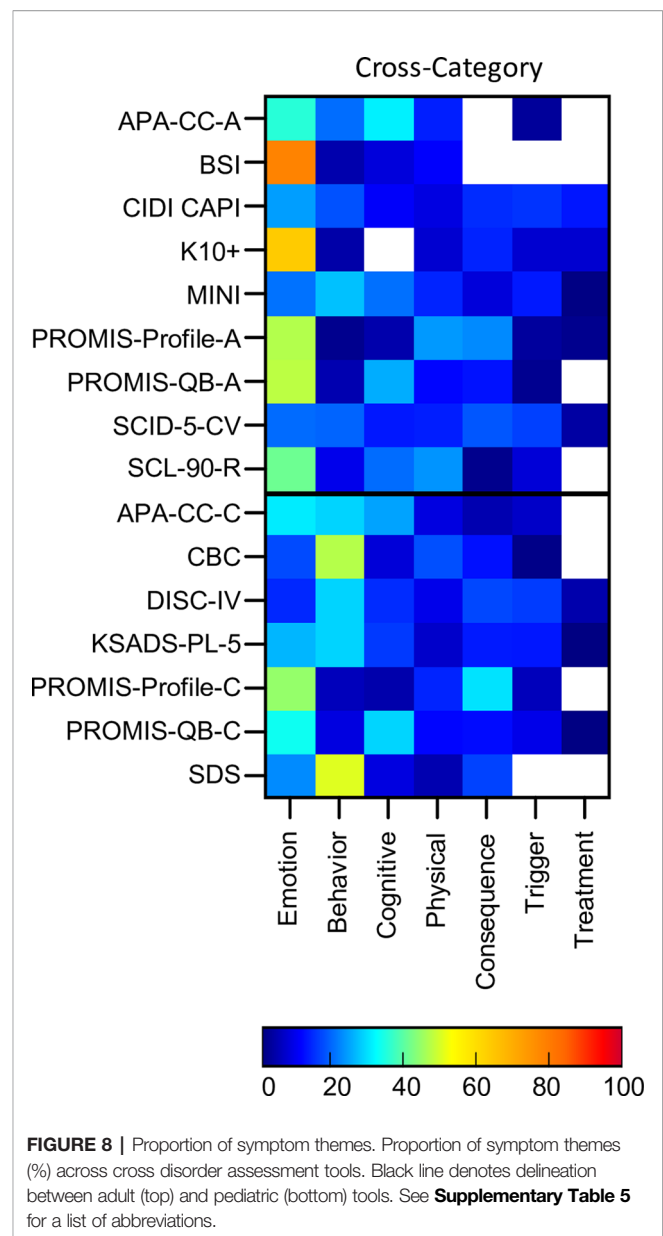
Cross Disorder Assessments

In addition to the assessment tools which were developed around the 10 disorder types, 16 cross-disorder tools were also included. These included nine assessments for adult populations and seven for pediatric populations (See **Supplementary Table 1** for details). Five of these tools were diagnostic and the remaining were auxiliary. For the modular clinical interviews and questionnaires (CIDI CAPI, DISC-IV, SCID-5, MINI, KSADS, PROMIS-QB-A, PROMIS-QB-C) only those modules which were related to the 10 disorders of interest were included.

Cross-disorder assessments, if comprehensive, should cover all symptoms and themes across these 10 disorders. Indeed 100% of symptoms were asked about across all tools in total (**Figure 4**). However, only 21% of symptoms were assessed by 80% or more of the assessment tools (Attention & Concentration & Mental Focus; Energy Levels; Pains & Aches; Attachment & Affiliation; Confidence & Self Judgement; Anger & Irritability; Fear, Panic & Anxiety; Mood & Outlook; Interest, Effort & Motivation). The maximum symptom categories covered by any single cross-disorder tool was 86% (SCID-5-CV) indicating that none of the available tools covered the full breadth of mental health symptoms corresponding to these 10 disorders.

In terms of symptom themes, many cross-disorder tools were biased in terms of symptom theme (**Figure 8, Supplementary Table 3**). For instance, the BSI and K10+ were primarily emotion focused (78% and 63% respectively) while only 14% of questions were emotion themed in others (DISC-IV). Others such as the CBC and SDS were more heavily weighted towards behavioural symptoms (47% and 50% respectively). The majority of tools asked about symptom triggers and consequences, while only half asked about symptom treatment.

On average, the cross-disorder tools were only 32% similar with similarity scores for individual pairs of assessments ranging from 9% (BSI and K10+) to 65% (KSADS and DISC-IV; see **Supplementary Table 4** for details). To determine whether there was any difference between diagnostic and auxiliary tools we computed similarity scores for each group separately as well between them. This showed that, on average, auxiliary tools were 29% similar, diagnostic tools were 58% similar, and diagnostic and auxiliary tools were 31% similar. Thus, the diagnostic interviews were considerably more alike compared to the auxiliary tools. Comparisons between adult and pediatric tools showed that adult tools were, on average, 32% similar, pediatric tools were 32% similar, and the similarity score between the adult



and pediatric tools was 33%, suggesting that the groups were comparable in terms of the similarity between tools.

DISCUSSION

Mental health determination is reliant, first and foremost, on the way that a patient's symptoms are assessed. The questions included on a questionnaire, or asked during an interview, determine the responses which are used within clinical and research domains to reveal psychological challenges, identify neurobiological disruptions or understand socio-environmental impacts. Here our analysis of 110 assessment tools spanning 10 different disorders, as well as 16 which took a cross-disorder perspective, revealed the high level of inconsistency in this

expansive assessment landscape which, in turn, has consequences for research and clinical reproducibility, as well as for the wider system of mental health diagnosis and evaluation. This inconsistency was apparent when comparing assessment tools across different disorders, as well as when comparing those intended to assess the same disorder. Significantly, the variability extended not just to how symptoms were being assessed, but to which symptoms were assessed.

An Inconsistent Approach to Mental Health Assessment

Our analysis of the similarity of symptoms between assessments of the same disorder revealed similarity scores ranging from 29% for bipolar disorder and schizophrenia to 58% for OCD. Between individual pairs of questionnaires assessing the same disorder, there were many cases where similarity of symptoms was lower than 20% (e.g., PCL-5 and SPRINT for PTSD; STAI and ZAS for anxiety, BSI and K10+ for cross-category). Thus, tools designed to assess the same disorder(s) showed considerable inconsistency in the symptoms that they were considering. Consequently, two experimental studies assessing patients with the same clinical diagnosis, but using different tools to assess symptom severity, may deliver different results because they are assessing a different set of symptoms. Furthermore, when combined with differences in the manner a symptom is assessed and the comorbid nature of symptom experience across individuals (140–142), the difficulties in searching for consistency and reproducibility are amplified.

While on one hand there were inconsistencies in symptom assessment within disorders, on the other hand there was also considerable overlap between disorders. Sixty percent of all mental health symptoms were included in the assessment of at least half of all disorders. Furthermore, 7% of all symptoms were assessed within all disorders, while a further 9% were assessed within all but one disorder. Similarity scores between disorders, which considered the differential weighting of symptoms between disorders, could be as high as 28% overall (depression and bipolar disorder) and 59% between two individual questionnaires (SANS, schizophrenia and SCQ, ASD, which similarly assess symptoms of Attachment & Affiliation; Interpersonal; Nonverbal Communication; Self-perception, Awareness & Appearance; Interest, Effort & Motivation). This high level of overlap in symptom assessment between disorders reflects the fact that the boundaries between mental health disorders are blurry, but also questions the relevance of “disorder-specific” assessment tools. Unlike most physiological disorders where symptoms are used as a guide to diagnosis of an underlying etiology (e.g., fever as a symptom of underlying viral or bacterial infection), mental health disorders are defined purely based on symptom sets and not any underlying physiology or causal factor. Thus, the definition of a disorder as a particular set of symptoms and the subsequent assessment of a “disorder” by a modified set of symptoms results in a circular logic or definitional contradiction. This introduces considerable confusion into the entire system of disorder classification and

highlights the challenge of achieving logical consistency in a system of disorders defined on the basis of symptoms rather than etiology or other causal factors.

Assessment tools also varied in terms of the symptom themes that they assessed, with tools for some disorders such as depression and anxiety being more heavily weighted towards asking about emotions and physical aspects of symptoms, while tools for disorders such as ADHD and addiction were more heavily weighted towards questions asking about behavior. This distinction corresponds with the common delineation between internalizing and externalizing spectra (143–145) and raises the question as to whether the delineation solely reflects the clinical manifestation of these disorders, or whether it is also constrained by the corresponding assessment tools. Furthermore, the distribution of symptom themes varied considerably within some disorders. For example, assessments of ASD typically asked about emotional, cognitive and behavioral symptoms to reflect the different manifestations of the disorder (146), but there was considerable heterogeneity in terms of which symptom theme was given the most weighting across different ASD assessments. For example, some assessments gave prominence to behavioral symptoms (e.g., ADI-R), others emphasized cognitive symptoms (all AQ assessment tools) and still others emphasized emotional symptoms (e.g., EQSQ). This bias towards different symptom themes could result in a different clinical picture emerging of the patient’s functioning depending on whether the assessment focused more on their emotional concerns or their behavioral ones. Similarly, when attempting to relate symptom experience against neurophysiological changes or socio-environmental factors, then, for example, assessment tools which highlight the severity of cognitive symptoms may produce different results compared to assessments highlighting behavioral symptoms.

Beyond which symptoms and what aspect of these symptoms were being assessed, the variation also extended to the way that symptoms were being assessed. In particular, the framing of the question in terms of whether it was asking about the frequency, severity, presence, duration, or timing of a symptom was inconsistently applied across assessment tools across disorders. Some disorders emphasized the frequency of symptoms (e.g., for anxiety) while others emphasized the severity of symptoms (e.g., PTSD, schizophrenia, OCD). Similarly, the time period over which symptoms were assessed also varied across disorders. Some disorders were dominated by assessments which considered only recent symptom experience, namely anxiety, bipolar disorder, and depression, while others were assessed by predominantly considering symptoms over the past few months (ADHD, PTSD, OCD) or years (addiction). One caveat to this is that some assessments have different versions which focus on different time periods (e.g., the past month version of CAPS -5 was included in this analysis, but there is also a past week version which was not) and the analysis was therefore somewhat influenced by this. However, the overall consequence of this variability is substantial noise in assessment that both hampers research and the assessment of clinical progress in a patient.

Our results are also in line with previous explorations of the heterogeneity of symptom assessment across different depression

tools, where Fried (15) concluded that the “substantial heterogeneity of the depressive syndrome and low overlap among scales may lead to research results idiosyncratic to particular scales used, posing a threat to the replicability and generalizability of depression research”. Given the inconsistency that we have reported here across a much larger group of assessment tools, it warrants the expansion of this conclusion beyond depression to cover the entirety of the mental health field.

Limitations

In interpreting this result there are several limitations to consider. Firstly, as with any qualitative approach, a degree of judgment and subjectivity is necessary to formulate a systematic coding method. In addition, our methodology involved a subjective grouping of symptoms based on our best judgement, but does not represent the only way, or a hard and fast way, of symptom grouping. A different group of researchers may consolidate symptoms into a different arrangement of symptom categories which could result in a different pattern of results, depending on the number of symptom groupings that they used. Nonetheless, we believe it would be unlikely to change the main finding of vast inconsistency.

Secondly, our analysis of mental health assessment tools was based on the semantic content of the questionnaires and interviews, rather than a patient’s response to those questions. Thus, while the final symptom categories used here are as semantically distinct as possible, they may not be independent from the perspective of underlying etiology in the same way as, for example, high fever and fatigue are semantically distinct but can arise from the same underlying cause.

Another limitation is that for a minority of assessment tools, there were a subset of questions which were included in the assessment, but not included in the assessment scoring. For example, the HCL-32 had 11 unscored questions, the WURS had 36 unscored questions and the YBOCS™ had 60 unscored questions from its checklist section. However, the inclusion of these additional unscored questions generally had only a small impact on the similarity scores as shown in the results. Furthermore, the inclusion of both diagnostic and auxiliary tools, did contribute to some of the differences, especially for the cross-disorder tools, but was not sufficient to explain the picture of heterogeneity across tools.

Finally, we note that our analysis covered 10 disorders selected based on their dominance in the mental health field, excluding less common disorders so as to limit the scope of this study. Furthermore, assessments of age-related disorders such as dementia were not included as they were considered to reflect a different aspect of brain health. It is therefore not possible to say whether the assessment of these disorders align with the results shown here. In addition, it was not possible to include every single assessment available for each of the 10 disorders of interest, and therefore there are others which are not included here. The consequence of this is that our reported scores of symptom similarity are not absolute, but rather generally representative of the field.

Implications for Mental Health Diagnosis and Research

Historically speaking, mental health diagnosis has been based around disorder-based diagnostic systems of DSM and ICD which map clusters of symptoms onto specific disorders based on phenomenological impressions of psychiatric dysfunction. This classical model of diagnosis has clinical utility in helping psychiatrists and clinical psychologists identify and respond to a diagnostic label corresponding to the array of emotional, cognitive, physical, and behavioral symptoms exhibited by the patient, that in turn guides a path of treatment. It also supports research linking socio-environmental and experiential factors to the onset and trajectory of a disorder. However, the effectiveness of this approach is heavily dependent on having a standardized assessment of symptoms, and clear boundaries between disorders. In contrast, the high degree of symptom overlap between disorders and the inconsistency within disorders shown here, raises concerns about whether disorder prevalence and patient outcomes are simply reflections of the choice of assessment tool. It also has implications for the development and delivery of patient-tailored cognitive-behavioral therapies and pharmacological interventions which target those psychological symptoms, where inconsistencies across symptom assessment may result in suboptimal treatment selection. In addition, such inconsistencies in symptom assessment can introduce challenges when making comparisons across research studies, thus weakening our understanding of psychopathological mechanisms and relevant risk factors. Furthermore, when combined with the comorbid nature of symptom presentation in patients, it exposes the immense challenges faced by psychiatric medicine today.

The heterogeneity of assessment tools also significantly hinders the search for disorder specific and transdiagnostic biomarkers and treatments where results of research studies and clinical trials may depend heavily on the particular bias of the selected assessment tool(s), and where physiological challenges or disturbances do not necessarily align with disorder demarcations. For example, criticism of the current categorical model of disorder classification (147, 148) has led to the emergence of a number of DSM-alternative initiatives that are disorder-independent or spectrum-based. For example, the Research Domain Criteria (RDoC) framework from the National Institute of Mental Health (NIMH) has been developed to provide an arena in which researchers can focus on the neurobiological and neurophysiological markers of psychiatric function and dysfunction, without needing to be aligned to a particular disorder taxonomy and provides a classification structure based around causal mechanisms at multiple levels of genetics, physiology and neuroscience based on data from animals and humans (149–151). In parallel, the Hierarchical Taxonomy of Psychopathology (HiTOP) takes learnings from the dimensional nature of psychopathology and is based on a hierarchical model of mental dysfunction, dominated by externalizing, internalizing and thought disorder spectra (152, 153). It is also affiliated with the development of a single transdiagnostic factor of psychopathology, or “p”, as a way of

encapsulating an individual's propensity to develop all forms of psychopathology (154). However, these models are, as yet, more challenging to operationalize within a clinical setting (155, 156) and the RDoC model has been criticized for specifically downplaying the importance of psychosocial factors (157). The emergence of these alternative transdiagnostic and disorder agnostic approaches are also not immune to inconsistencies in symptom assessment, highlighting the need for tools which have greater standardization in symptom assessment across the mental health landscape, strengthening comparisons across studies and limiting biases and inconsistencies invoked by choosing different sets of assessment tools.

Furthermore, the definition of what actually denotes a mental disorder, and where the boundary between normality and disorder really lies is challenging within such a categorical assessment framework (158). With many "symptoms" such as sadness, anxiousness, and risk-taking representing normal human behaviors, the quantitative boundary between normal and disorder is an important one. Therefore, tools and the research around them, must also strive to understand normality as a way to understand abnormality (159).

The Potential of Cross-Disorder Assessment

Given the overlap of disorders and the realization that comorbidity is the norm, rather than the exception (160, 161), there is now also increased traction behind the search for transdiagnostic markers of psychiatric dysfunction (11, 162, 163). This is typically done by looking across groups of patients with differential disorder diagnoses, rather than taking a disorder agnostic approach. Disruptions in brain structure and function, cognition and behavior have all been explored from a transdiagnostic perspective, revealing cross-disorder commonalities in the disruption of resting-state functional connectivity (164), cognitive control (165), reward responsivity (166), gray matter volume (167), and social cognition (168). Such work is also supported by evidence showing that some discretely labeled disorders have shared genetic etiology (169–171). Many transdiagnostic studies rely on the DSM system as the diagnostic gold standard (172) despite the heterogeneity in diagnostic symptom criteria both within and across the disorder specific modules of the DSM (173), in line with the results shown here on a larger scale. Furthermore, as we have demonstrated here, none of the available cross-disorder tools provide a comprehensive coverage of the symptoms across all of the 10 disorders of interest. The one with the widest coverage of symptoms was SCID-5-CV at only 86%. Furthermore, the cross-disorder tools themselves were highly variable (especially the auxiliary ones) with an overall average symptom similarity of only 32%. The development of transdiagnostic or disorder agnostic methods that provide a standardized assessment spanning the full breadth of psychiatric symptoms could therefore help advance these endeavors.

While transdiagnostic approaches typically focus on making comparisons at a disorder level, other approaches have focused their efforts at the symptom level, by modelling the interrelationship

between co-occurring symptoms and the evolution of symptom covariance using network structures both within disorders and across disorders (174–177). This approach represents individual symptoms as nodes and hubs within a network where the interconnecting lines depict the strength of interrelation between symptoms. The flexibility of this approach, where the networks can adapt and shift both over time to reflect clinical staging, and across individuals, provides an opportunity for delivering a more personalized approach to clinical diagnosis and treatment which is tailored to the individual needs of the patient based on their unique array of symptoms, rather than relying on a "best fit" scenario, as is currently the case. As the network structure is based on reported symptoms, it is therefore also heavily dependent on having standardized assessment tools that can be used to obtain this information.

Conclusion

In sum, the inevitable evolution of mental health research, diagnosis and treatment into a framework which is closer to underlying biology, transdiagnostic factors or symptom networks, would benefit from a new suite of standardized assessment tools that leave behind historical clustering of symptom criteria and instead provide a disorder agnostic perspective that spans the breadth of mental health function and dysfunction. Like others (178), we join the call for the development of such tools, proposing that attention be given to standardizing not just the suite of symptoms assessed but also to the various aspects of symptom theme, framing and quantification of symptom timing and severity.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

JN and DH performed the analysis and discussed the results with TT. JN drafted the manuscript. TT and JN revised the manuscript. JN, DH, and TT approved the final version, and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by internal funding from Sapien Labs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsy.2020.00076/full#supplementary-material>

REFERENCES

1. APA. (2013). *Diagnostic and statistical manual of mental disorders, 5th ed.* (Arlington, VA: American Psychiatric Association).
2. WHO. (2018). *International statistical classification of diseases and related health problems (11th Revision)* (WHO).
3. First M, Williams J, Karg R, Spitzer R. *Structured Clinical Interview for DSM-5 Disorders, Clinician Version (SCID-5-CV)*. Arlington, VA: American Psychiatric Association (2016).
4. Spitzer RL, Kroenke K, Williams JBW and the Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* (1999) 282(18):1737–44. doi: 10.1001/jama.282.18.1737
5. Santor DA, Gregus M, Welch A. Eight decades of measurement in depression. *Meas: Interdiscip Res Perspect* (2006) 4(3):135–55. doi: 10.1207/s15366359mea0403_1
6. Wisco BE, Miller MW, Wolf EJ, Kilpatrick D, Resnick HS, Badour CL, et al. The impact of proposed changes to ICD-11 on estimates of PTSD prevalence and comorbidity. *Psychiatry Res* (2016) 240:226–33. doi: 10.1016/j.psychres.2016.04.043
7. Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* (2017) 20: (3) 365–77. doi: 10.1038/nn.4478
8. Insel TR, Quirion R. Psychiatry as a clinical neuroscience discipline. *JAMA* (2005) 294(17):2221–4. doi: 10.1001/jama.294.17.2221
9. Venkatasubramanian G, Keshavan MS. Biomarkers in psychiatry - a critique. *Ann Neurosci* (2016) 23(1):3–5. doi: 10.1159/000443549
10. Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. *Nat Med* (2016) 22:1248. doi: 10.1038/nm.4190
11. Pinto JV, Moulin TC, Amaral OB. On the transdiagnostic nature of peripheral biomarkers in major psychiatric disorders: a systematic review. *Neurosci Biobehav Rev* (2017) 83:97–108. doi: 10.1016/j.neubiorev.2017.10.001
12. McLoughlin G, Makeig S, Tsuang MT. In search of biomarkers in psychiatry: EEG-based measures of brain function. *Am J Med Genet Part B: Neuropsychiatr Genet* (2014) 165(2):111–21. doi: 10.1002/ajmg.b.32208
13. Banner NF. Mental disorders are not brain disorders. *J Eval Clin Pract* 19 (2013) (3):509–13. doi: 10.1111/jep.12048
14. Borsboom D, Cramer AJO, Kalis A. Brain disorders? Not really: why network structures block reductionism in psychopathology research. *Behav Brain Sci* (2019) 42:e2. doi: 10.1017/S0140525X17002266
15. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *J Affect Disord* (2017) 208:191–7. doi: 10.1016/j.jad.2016.10.019
16. Newson JJ, Thiagarajan TC. EEG frequency bands in psychiatric disorders: a review of resting state studies. *Front Hum Neurosci* (2019) 12:521. doi: 10.3389/fnhum.2018.00521
17. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation (1996).
18. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psychol Meas* (1977) 1(3):385–401. doi: 10.1177/014662167700100306
19. Eaton WW MC, Smith C, Tien A, Ybarra M. Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R). *The use of Psychological testing for treatment planning and outcomes assessment: Instruments for adults, 3rd ed.* Vol. 3. . Mahwah, NJ: Lawrence Erlbaum Associates Publishers (2004) p. 363–77.
20. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry* (1987) 150(6):782–6. doi: 10.1192/bjp.150.6.782
21. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* (1982) 17(1):37–49. doi: 10.1016/0022-3956(82)90033-4
22. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin Gerontol* (1986) 5(1-2):165–73. doi: 10.1300/J018v05n01_09
23. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* (1960) 23(1):56. doi: 10.1136/jnnp.23.1.56
24. Rush JA, Giles DE, Schlessler MA, Fulton CL, Weissenburger J, Burns C. The inventory for depressive symptomatology (IDS): preliminary findings. *Psychiatry Res* (1986) 18(1):65–87. doi: 10.1016/0165-1781(86)90060-0
25. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* (1979) 134(4):382–9. doi: 10.1192/bjp.134.4.382
26. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* (2003) 54(5):573–83. doi: 10.1016/S0006-3223(02)01866-8
27. Zung WWK. A self-rating depression scale. *JAMA Psychiatry* (1965) 12 (1):63–70. doi: 10.1001/archpsyc.1965.01720310065008
28. Kovacs M. *Children's Depression Inventory 2 (CDI 2)*. 2nd ed. North Tonawanda, NY: Multi-Health Systems Inc (2011).
29. Kovacs M. The Children's Depression, Inventory (CDI). *Psychopharmacol Bull* (1985) 21(4):995–8.
30. Weissman MM, Orvaschel H, Padian N. Children's symptom and social functioning self-report scales comparison of mothers' and children's reports. *J Nerv Ment Dis* (1980) 168(12):736–40. (New Jersey: John Wiley & Sons). doi: 10.1097/00005053-198012000-00005
31. Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int J Methods Psychiatr Res* (1995) 5(4):237–49.
32. Reynolds WM. Reynolds Adolescent Depression Scale. In: Hilsenroth MJ, Segal DL, Hersen M, editors. *Comprehensive handbook of Psychological assessment, Volume 2: Personality assessment and psychopathology, 2nd Edition*. New Jersey: John Wiley & Sons (2004). p. 224–36.
33. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* (1988) 56(6):893–7. doi: 10.1037/0022-006X.56.6.893
34. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *JAMA Internal Med* (2006) 166 (10):1092–7. doi: 10.1001/archinte.166.10.1092
35. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* (1959) 32(1):50–5. doi: 10.1111/j.2044-8341.1959.tb00467.x
36. Fresco DM, Coles ME, Heimberg RG, Liebowitz MR, Hami S, Stein MB, et al. The liebowitz social anxiety scale: a comparison of the psychometric properties of self-report and clinician-administered formats. *Psychol Med* (2001) 31(6):1025–35. doi: 10.1017/S0033291701004056
37. Liebowitz MR. Social Phobia. In: Klein DF, editor. *Anxiety Mod Trends Pharmacopsychiatry* 22. Basel: Karger (1987). p. 141–73.
38. Connor KM, Davidson JRT, Churchill LE, Sherwood A, Weisler RH, Foa E. Psychometric properties of the Social Phobia Inventory (SPIN): new self-rating scale. *Br J Psychiatry* (2000) 176(4):379–86. doi: 10.1192/bjp.176.4.379
39. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press (1983).
40. Zung WWK. A rating instrument for anxiety disorders. *Psychosomatics* (1971) 12(6):371–9. doi: 10.1016/S0033-3182(71)71479-0
41. March J. *Multi-dimensional anxiety scale for children*. North Tonawanda, NY: Multi Health Systems (1998).
42. March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. *J Am Acad Child Adolesc Psychiatry* (1997) 36 (4):554–65. doi: 10.1097/00004583-199704000-00019
43. Reynolds CR, Richmond BO. What I think and feel: a revised measure of children's manifest anxiety. *J Abnormal Child Psychol* (1978) 6(2):271–80. doi: 10.1007/BF00919131
44. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry* (1997) 36(4):545–53. doi: 10.1097/00004583-199704000-00018
45. Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *J Am Acad Child*

- Adolesc Psychiatry* (1999) 38(10):1230–6. doi: 10.1097/00004583-199910000-00011
46. Spence SH. A measure of anxiety symptoms among children. *Behav Res Ther* (1998) 36(5):545–66. doi: 10.1016/S0005-7967(98)00034-5
 47. Weathers FW, Bovin MJ, Lee DJ, Sloan DM, Schnurr PP, Kaloupek DG, et al. The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): development and initial psychometric evaluation in military veterans. *Psychol Assess* (2018) 30(3):383–95. doi: 10.1037/pas0000486
 48. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. *J Gen Intern Med* (2016) 31(10):1206–11. doi: 10.1007/s11606-016-3703-5
 49. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. *J Traumatic Stress* (2015) 28(6):489–98. doi: 10.1002/jts.22059
 50. Foa EB, Cashman L, Jaycox L, Perry K. The validation of a self-report measure of posttraumatic stress disorder: the posttraumatic diagnostic scale. *Psychol Assess* (1997) 9(4):445–51. doi: 10.1037/1040-3590.9.4.445
 51. Connor KM, Davidson JRT. SPRINT: a brief global assessment of post-traumatic stress disorder. *Int Clin Psychopharmacol* (2001) 16(5):279–84. doi: 10.1097/00004850-200109000-00005
 52. Pynoos RS, Weathers FW, Steinberg AM, Marx BP, Layne CM, Kaloupek DG, et al. (2015). Clinician-administered PTSD scale for DSM-5 - child/adolescent version. Scale available from the National Center for PTSD at www.ptsd.va.gov.
 53. Foa EB, Asnaani A, Zang Y, Capaldi S, Yeh R. Psychometrics of the child PTSD symptom scale for DSM-5 for trauma-exposed children and adolescents. *J Clin Child Adolesc Psychol* (2018) 47(1):38–46. doi: 10.1080/15374416.2017.1350962
 54. Altman EG, Hedeker D, Peterson JL, Davis JM. The altman self-rating mania scale. *Biol Psychiatry* (1997) 42(10):948–55. doi: 10.1016/S0006-3223(96)00548-3
 55. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: Towards a self-assessment tool for hypomanic symptoms in outpatients. *J Affect Disord* (2005) 88(2):217–33. doi: 10.1016/j.jad.2005.05.011
 56. Bauer MS, Crits-Christoph P, Ball WA, Dewees E, McAllister T, Alahi P, et al. Independent assessment of manic and depressive symptoms by self-rating: scale characteristics and implications for the study of mania. *JAMA Psychiatry* (1991) 48(9):807–12. doi: 10.1001/archpsyc.1991.01810330031005
 57. Hirschfeld RMA, Williams JBW, Spitzer RL, Calabrese JR, Flynn L, Keck PE, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire. *Am J Psychiatry* (2000) 157(11):1873–5. doi: 10.1176/appi.ajp.157.11.1873
 58. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* (1978) 133(5):429–35. doi: 10.1192/bjp.133.5.429
 59. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the dimensional obsessive-compulsive scale. *Psychol Assess* (2010) 22(1):180–98. doi: 10.1037/a0018260
 60. Storch EA, Bagner D, Merlo LJ, Shapira NA, Geffken GR, Murphy TK, et al. Florida obsessive-compulsive inventory: development, reliability, and validity. *J Clin Psychol* (2007) 63(9):851–9. doi: 10.1002/jclp.20382
 61. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The obsessive-compulsive inventory: development and validation of a short version. *Psychol Assess* (2002) 14(4):485–96. doi: 10.1037/1040-3590.14.4.485
 62. Burns GL, Keortge SG, Formea GM, Sternberger LG. Revision of the Padua Inventory of obsessive compulsive disorder symptoms: distinctions between worry, obsessions, and compulsions. *Behav Res Ther* (1996) 34(2):163–73. doi: 10.1016/0005-7967(95)00035-6
 63. Thordarson DS, Radomsky AS, Rachman S, Shafran R, Sawchuk CN, Ralph Hakstian A. The Vancouver Obsessional Compulsive Inventory (VOCI). *Behav Res Ther* (2004) 42(11):1289–314. doi: 10.1016/j.brat.2003.08.007
 64. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The yale-brown obsessive compulsive scale: I. development, use, and reliability. *Arch Gen Psychiatry* (1989) 46(11):1006–11. doi: 10.1001/archpsyc.1989.01810110048007
 65. Storch EA, Rasmussen SA, Price LH, Larson MJ, Murphy TK, Goodman WK. Development and psychometric evaluation of the yale-brown obsessive-compulsive scale—second edition. *Psychol Assess* (2010) 22(2):223–32. doi: 10.1037/a0018492
 66. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown obsessive compulsive scale: reliability and validity. *J Am Acad Child Adolesc Psychiatry* (1997) 36(6):844–52. doi: 10.1097/00004583-199706000-00023
 67. Foa EB, Coles M, Huppert JD, Pasupuleti RV, Franklin ME, March J. Development and validation of a child version of the obsessive compulsive inventory. *Behav Ther* (2010) 41(1):121–32. doi: 10.1016/j.beth.2009.02.001
 68. Skinner HA, Horn JL. *Alcohol Dependence Scale (ADS): User's Guide*. Toronto, Canada: Addiction Research Foundation (1984).
 69. Skinner HA, Allen BA. Alcohol dependence syndrome: measurement and validation. *J Abnormal Psychol* (1982) 91(3):199–209. doi: 10.1037/0021-843X.91.3.199
 70. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. The fifth edition of the addiction severity index. *J Subst Abuse Treat* (1992) 9(3):199–213. doi: 10.1016/0740-5472(92)90062-S
 71. WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. *Addiction* (2002) 97(9):1183–94. doi: 10.1046/j.1360-0443.2002.00185.x
 72. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO Collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction* (1993) 88(6):791–804. doi: 10.1111/j.1360-0443.1993.tb02093.x
 73. Ewing JA. Detecting alcoholism: the CAGE questionnaire. *JAMA* (1984) 252(14):1905–7. doi: 10.1001/jama.1984.03350140051025
 74. Skinner HA. The drug abuse screening test. *Addictive Behav* (1982) 7(4):363–71. doi: 10.1016/0306-4603(82)90005-3
 75. Berman AH, Bergman H, Palmstierna T, Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a swedish population sample. *Eur Addict Res* (2005) 11(1):22–31. doi: 10.1159/000081413
 76. Selzer ML. The michigan alcoholism screening test: the quest for a new diagnostic instrument. *Am J Psychiatry* (1971) 127(12):1653–8. doi: 10.1176/ajp.127.12.1653
 77. Anton RF, Moak DH, Latham P. The obsessive compulsive drinking scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. *Alcohol: Clin Exp Res* (1995) 19(1):92–9. doi: 10.1111/j.1530-0277.1995.tb01475.x
 78. Selzer ML, Vinokur A, van Rooijen L. A self-administered Short Michigan Alcoholism Screening Test (SMAST). *J Stud Alcohol* (1975) 36(1):117–26. doi: 10.15288/jsa.1975.36.117
 79. Russell M. New assessment tools for risk drinking during pregnancy: T-ACE, TWEAK, and others. *Alcohol Health Res World* (1994) 18(1):55–61.
 80. Knight JR, Shrier LA, Bravender TD, Farrell M, Vander Bilt J, Shaffer HJ. A new brief screen for adolescent substance abuse. *JAMA Pediatr* (1999) 153(6):591–6. doi: 10.1001/archpedi.153.6.591
 81. Baron-Cohen S, Wheelwright S, Robinson J, Woodbury-Smith M. The Adult Asperger Assessment (AAA): a diagnostic method. *J Autism Dev Disord* (2005) 35(6):807. doi: 10.1007/s10803-005-0026-5
 82. Allison C, Auyeung B, Baron-Cohen S. Toward Brief “Red Flags” for Autism Screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *J Am Acad Child Adolesc Psychiatry* (2012) 51(2):202–12.e7. doi: 10.1016/j.jaac.2011.11.003
 83. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The Autism-Spectrum Quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* (2001) 31(1):5–17. doi: 10.1023/A:1005653411471
 84. Hurlley RSE, Losh M, Parlier M, Reznick JS, Piven J. The broad autism phenotype questionnaire. *J Autism Dev Disord* (2007) 37(9):1679–90. doi: 10.1007/s10803-006-0299-3
 85. Baron-Cohen S, Wheelwright S. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord* (2004) 34(2):163–75. doi: 10.1023/B:JADD.0000022607.19833.00

86. Frith U, Hill EL, Baron-Cohen S, Richler J, Bisarya D, Gurusathyan N, et al. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos Trans R Soc London Ser B: Biol Sci* (2003) 358(1430):361–74. doi: 10.1098/rstb.2002.1206
87. Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* (1994) 24(5):659–85. doi: 10.1007/BF02172145
88. Rutter M, Le Couteur A, Lord C. *Autism Diagnostic Interview - Revised: Manual*. Los Angeles, CA: Western Psychological Services (2003).
89. Baron-Cohen S, Hoekstra RA, Knickmeyer R, Wheelwright S. The Autism-Spectrum Quotient (AQ)—adolescent version. *J Autism Dev Disord* (2006) 36(3):343. doi: 10.1007/s10803-006-0073-6
90. Auyeung B, Baron-Cohen S, Wheelwright S, Allison C. The autism spectrum quotient: children's version (AQ-Child). *J Autism Dev Disord* (2008) 38(7):1230–40. doi: 10.1007/s10803-007-0504-z
91. Ehlers S, Gillberg C, Wing L. A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *J Autism Dev Disord* (1999) 29(2):129–41. doi: 10.1023/A:1023040610384
92. Schopler E, Van Bourgondien ME, Wellman GJ, Love SR. *The Childhood Autism Rating Scale, Second Edition (CARS-2)*. Torrance, CA: Western Psychological Services (2010).
93. Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J Autism Dev Disord* (1980) 10(1):91–103. doi: 10.1007/BF02408436
94. Scott FJ, Baron-Cohen S, Bolton P, Brayne C. The CAST (Childhood Asperger Syndrome Test): preliminary development of a UK Screen for Mainstream primary-school-age children. *Autism* (2002) 6(1):9–31. doi: 10.1177/1362361302006001003
95. Auyeung B, Allison C, Wheelwright S, Baron-Cohen S. Brief report: development of the adolescent empathy and systemizing quotients. *J Autism Dev Disord* (2012) 42(10):2225–35. doi: 10.1007/s10803-012-1454-7
96. Auyeung B, Wheelwright S, Allison C, Atkinson M, Samarawickrema N, Baron-Cohen S. The children's empathy quotient and systemizing quotient: sex differences in typical development and in autism spectrum conditions. *J Autism Dev Disord* (2009) 39(11):1509. doi: 10.1007/s10803-009-0772-x
97. Gilliam JE. *Gilliam Autism Rating Scale-Third Edition*. Austin, TX: PRO-ED (2014).
98. Robins DL, Fein D, Barton ML, Green JA. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord* (2001) 31(2):131–44. doi: 10.1023/A:1010738829569
99. Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics* (2014) 133(1):37. doi: 10.1542/peds.2013-1813
100. Rutter M, Bailey A CL. *The Social Communication Questionnaire: Manual*. Los Angeles, CA: Western Psychological Services (2003).
101. Constantino JN GC. *Social Responsiveness Scale, Second Edition (SRS-2)*. Los Angeles, CA: Western Psychological Services (2012).
102. Ustun B, Adler LA, Rudin C, Faraone SV, Spencer TJ, Berglund P, et al. The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *JAMA Psychiatry* (2017) 74(5):520–6. doi: 10.1001/jamapsychiatry.2017.0298
103. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi EVA, et al. The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol Med* (2005) 35(2):245–56. doi: 10.1017/S0033291704002892
104. Conners C, Erhardt D, Sparrow E. *Conners' adult ADHD rating scales (CAARS)*. New York, NY: Multi-Health Systems, Inc. (1999).
105. Kooij JJS. *Adult ADHD Diagnostic assessment and treatment. 3rd Edition*. London: Springer-Verlag Publishing (2012). doi: 10.1007/978-1-4471-4138-9
106. Ward MF, Wender PH, Reimherr FW. The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *Am J Psychiatry* (1993) 150(6):885–90. doi: 10.1176/ajp.150.6.885
107. Conners C. *Conners 3rd edition manual*. Toronto, Ontario, Canada: Multi-Health Systems (2008).
108. Pelham WEJ, Gnagy EM, Greenslade KE, Milich R. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *J Am Acad Child Adolesc Psychiatry* (1992) 31(2):210–8. doi: 10.1097/00004583-199203000-00006
109. Wolraich ML, Lambert W, Doffing MA, Bickman L, Simmons T, Worley K. Psychometric properties of the vanderbilt ADHD diagnostic parent rating scale in a referred population. *J Pediatr Psychol* (2003) 28(8):559–68. doi: 10.1093/jpepsy/jsg046
110. Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Rep* (1962) 10(3):799–812. doi: 10.2466/pr0.1962.10.3.799
111. Ventura J, Lukoff D, Nuechterlein K, Liberman RP, Green MF, Shaner A. Brief psychiatric rating scale expanded version 4.0: scales anchor points and administration manual. *Int J Methods Psychiatr Res* (1993) 13:221–44.
112. Haro JM, Kamath SA, Ochoa S, Novick D, Rele K, Fargas A, et al. The clinical global impression-schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatrica Scand* (2003) 107(s416):16–23. doi: 10.1034/j.1600-0447.107.s416.5.x
113. Axelrod BN, Goldman RS, Alphas LD. Validation of the 16-item negative symptom assessment. *J Psychiatr Res* (1993) 27(3):253–8. doi: 10.1016/0022-3956(93)90036-2
114. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for Schizophrenia. *Schizophr Bull* (1987) 13(2):261–76. doi: 10.1093/schbul/13.2.261
115. Andreasen NC. The scale for the assessment of negative symptoms (sands): conceptual and theoretical foundations. *Br J Psychiatry* (1989) 155(S7):49–52. doi: 10.1192/S0007125000291496
116. Andreasen NC. Negative symptoms in Schizophrenia: definition and reliability. *JAMA Psychiatry* (1982) 39(7):784–8. doi: 10.1001/archpsyc.1982.04290070020005
117. Andreasen NC. *The Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, Iowa: The University of Iowa (1983).
118. Andreasen NC. *The Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, Iowa: The University of Iowa (1984).
119. Henderson M, Freeman CPL. A self-rating scale for bulimia the 'BITE'. *Br J Psychiatry* (1987) 150(1):18–24. doi: 10.1192/bjp.150.1.18
120. Garner DM, Olmsted MP, Bohr Y, Garfinkel PE. The eating attitudes test: psychometric features and clinical correlates. *Psychol Med* (1982) 12(4):871–8. doi: 10.1017/S0033291700049163
121. Stice E, Telch CF, Rizvi SL. Development and validation of the eating disorder diagnostic scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychol Assess* (2000) 12(2):123–31. doi: 10.1037/1040-3590.12.2.123
122. Fairburn C, Cooper Z, O'Connor M. Eating Disorder Examination. In: Fairburn CG, editor. *Cognitive Behaviour Therapy and Eating Disorders*. New York: Guildford Press (2008).
123. Garner DM. *Eating Disorder Inventory-3. Professional Manual*. Lutz F, editor. Florida: Psychological Assessment Resources, Inc. (2004).
124. Morgan JF, Reid F, Lacey JH. The SCOFF questionnaire: assessment of a new screening tool for eating disorders. *BMJ* (1999) 319(7223):1467. doi: 10.1136/bmj.319.7223.1467
125. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. *Psychol Med* (1983) 13(3):595–605. doi: 10.1017/S0033291700048017
126. Andrews G, Peters L. The psychometric properties of the composite international diagnostic interview. *Soc Psychiatry Psychiatr Epidemiol* (1998) 33(2):80–8. doi: 10.1007/s001270050026
127. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The composite international diagnostic interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *JAMA Psychiatry* (1988) 45(12):1069–77. doi: 10.1001/archpsyc.1988.01800360017003
128. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific Psychological distress. *Psychol Med* (2002) 32(6):959–76. doi: 10.1017/S0033291702006074
129. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and

- validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* (1998) 59(Suppl 20):22–33.
130. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* (2010) 63(11):1179–94. doi: 10.1016/j.jclinepi.2010.04.011
 131. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an nih roadmap cooperative group during its first two years. *Med Care* (2007) 45(5):S3–S11. (Towson, MD: Clinical Psychometric Research). doi: 10.1097/01.mlr.0000258615.42478.55
 132. Derogatis L. *SCL-90-R: Administration, Scoring of Procedures Manual-II for the R (revised) Version and Other Instruments of the Psychopathology Rating Scale Series*. (Towson, MD: Clinical Psychometric Research) (1992).
 133. Achenbach TM, Rescorla LA. *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families (2001).
 134. Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* (2000) 39(1):28–38. doi: 10.1097/00004583-200001000-00014
 135. Kaufman J, Birmaher, B, Brent, D, Rao, UMA, Flynn, C, Moreci, P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Psy* (1997) 36(7):980–8. doi: 10.1097/00004583-199707000-00021
 136. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* (2001) 40(11):1337–45. doi: 10.1097/00004583-200111000-00015
 137. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behav Res Ther* (1997) 35(1):79–89. doi: 10.1016/S0005-7967(96)00068-X
 138. Lovibond SH, Lovibond PF. *Manual for the Depression Anxiety Stress Scales*. 2nd. Ed. Sydney: Psychology Foundation (1995).
 139. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scand* (1983) 67(6):361–70. doi: 10.1111/j.1600-0447.1983.tb09716.x
 140. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. *J Affect Disord* (2015) 172:96–102. doi: 10.1016/j.jad.2014.10.010
 141. Zoellner LA, Pruitt LD, Farach FJ, Jun JJ. Understanding heterogeneity in PTSD: Fear, Dysphoria, and Distress. *Depression Anxiety* (2014) 31(2):97–106. doi: 10.1002/da.22133
 142. Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry* (2018) 75(11):1146–55. doi: 10.1001/jamapsychiatry.2018.2467
 143. Markon KE. Modeling psychopathology structure: a symptom-level analysis of Axis I and II disorders. *Psychol Med* (2010) 40(2):273–88. doi:10.1017/S0033291709990183.
 144. Krueger RF. The Structure of common mental disorders. *JAMA Psychiatry* (1999) 56(10):921–6. doi: 10.1001/archpsyc.56.10.921
 145. Krueger RF, Eaton NR. Transdiagnostic factors of mental disorders. *World Psychiatry* (2015) 14(1):27–9. doi: 10.1002/wps.20175
 146. Masi A, DeMayo MM, Glozier N, Guastella AJ. An overview of Autism spectrum disorder, heterogeneity and treatment options. *Neurosci Bull* (2017) 33(2):183–93. doi: 10.1007/s12264-017-0100-y
 147. Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, Reed GM. Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychol Sci Public Interest* (2017) 18(2):72–145. doi: 10.1177/1529100617727266
 148. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* (2012) 17:1174. doi: 10.1038/mp.2012.105
 149. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* (2010) 167(7):748–51. doi: 10.1176/appi.ajp.2010.09091379
 150. Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry* (2014) 171(4):395–7. doi: 10.1176/appi.ajp.2014.14020138
 151. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* (2013) 11(1):126. doi: 10.1186/1741-7015-11-126
 152. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J Abnormal Psychol* (2017) 126(4):454–77. doi: 10.1037/abn0000258
 153. Conway CC, Forbes MK, Forbush KT, Fried EI, Hallquist MN, Kotov R, et al. A hierarchical taxonomy of psychopathology can transform mental health research. *Perspect Psychol Sci* (2019) 14(3):419–36. doi: 10.1177/1745691618810696
 154. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p Factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* (2013) 2(2):119–37. doi: 10.1177/2167702613497473
 155. First M. Commentary on Krueger and Eaton's "Personality traits and the classification of mental disorders: toward a more complete integration in DSM-5 and an empirical model of psychopathology": real-world considerations in implementing an empirically based dimensional model of personality in DSM-5. *Personal Disord* (2010) 1:123–6. doi: 10.1037/a0019975
 156. Akram F, Giordano J. Research domain criteria as psychiatric nosology: conceptual, practical and neuroethical implications. *Cambridge Q Healthcare Ethics* (2017) 26(4):592–601. doi: 10.1017/S096318011700010X
 157. Lilienfeld SO, Treadway MT. Clashing diagnostic approaches: DSM-ICD Versus RDoC. *Annu Rev Clin Psychol* (2016) 12(1):435–63. doi: 10.1146/annurev-clinpsy-021815-093122
 158. Wakefield JC. Diagnostic issues and controversies in DSM-5: return of the false positives problem. *Annu Rev Clin Psychol* (2016) 12(1):105–32. doi: 10.1146/annurev-clinpsy-032814-112800
 159. Nesse RM, Stein DJ. Towards a genuinely medical model for psychiatric nosology. *BMC Med* (2012) 10(1):5. doi: 10.1186/1741-7015-10-5
 160. Maj M. 'Psychiatric comorbidity': an artefact of current diagnostic systems? *Br J Psychiatry* (2005) 186(3):182–4. doi: 10.1192/bjp.186.3.182
 161. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, et al. Exploring comorbidity within mental disorders among a danish national population. *JAMA Psychiatry* (2019) 76(3):259–70. doi: 10.1001/jamapsychiatry.2018.3658
 162. Etkin A, Cuthbert B. Beyond the DSM: development of a transdiagnostic psychiatric neuroscience course. *Acad Psychiatry* (2014) 38(2):145–50. doi: 10.1007/s40596-013-0032-4
 163. Williams LM, Goldstein-Piekariski AN, Chowdhry N, Grisanzio KA, Haug NA, Samara Z, et al. Developing a clinical translational neuroscience taxonomy for anxiety and mood disorder: protocol for the baseline-follow up Research domain criteria Anxiety and Depression ("RAD") project. *BMC Psychiatry* (2016) 16(1):68. doi: 10.1186/s12888-016-0771-3
 164. Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* (2018) 9(1):3003. doi: 10.1038/s41467-018-05317-y
 165. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *Am J Psychiatry* (2017) 174(7):676–85. doi: 10.1176/appi.ajp.2017.16040400
 166. Sharma A, Wolf DH, Ciric R, Kable JW, Moore TM, Vandekar SN, et al. Common dimensional reward deficits across mood and psychotic disorders: a connectome-wide association study. *Am J Psychiatry* (2017) 174(7):657–66. doi: 10.1176/appi.ajp.2016.16070774
 167. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* (2015) 72(4):305–15. doi: 10.1001/jamapsychiatry.2014.2206
 168. Cotter J, Granger K, Backx R, Hobbs M, Looi CY, Barnett JH. Social cognitive dysfunction as a clinical marker: a systematic review of meta-analyses across

- 30 clinical conditions. *Neurosci Biobehav Rev* (2018) 84:92–9. doi: 10.1016/j.neubiorev.2017.11.014
169. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* (2013) 45:984. doi: 10.1038/ng.2711
170. Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat Neurosci* (2016) 19:1392. doi: 10.1038/nn.4411
171. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G, et al. Psychiatric genomics: an update and an agenda. *Am J Psychiatry* (2017) 175(1):15–27. doi: 10.1176/appi.ajp.2017.17030283
172. Fusar-Poli P, Solmi M, Brondino N, Davies C, Chae C, Politi P, et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry* (2019) 18(2):192–207. doi: 10.1002/wps.20631
173. Allsopp K, Read J, Corcoran R, Kinderman P. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Res* (2019) 279:15–22. doi: 10.1016/j.psychres.2019.07.005
174. Borsboom D, Cramer AOJ. Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol* (2013) 9(1):91–121. doi: 10.1146/annurev-clinpsy-050212-185608
175. Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. comorbidity: a network perspective. *Behav Brain Sci* (2010) 33(2-3):137–50. doi: 10.1017/S0140525X09991567
176. Boschloo L, van Borkulo CD, Rhemtulla M, Keyes KM, Borsboom D, Schoevers RA. The network structure of symptoms of the diagnostic and statistical manual of mental disorders. *PLoS One* (2015) 10(9):e0137621. doi: 10.1371/journal.pone.0137621
177. Cramer AOJ, van Borkulo CD, Giltay EJ, van der Maas HLJ, Kendler KS, Scheffer M, et al. Major depression as a complex dynamic system. *PLoS One* (2016) 11(12):e0167490. doi: 10.1371/journal.pone.0167490
178. Van Heugten Van Der Kloet D, van Heugten T. The classification of psychiatric disorders according to DSM-5 deserves an internationally standardized Psychological test battery on symptom level. *Front Psychol* (2015) 6:1108. doi: 10.3389/fpsyg.2015.01108

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Newson, Hunter and Thiagarajan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.