
PHQ-9, Life Impact, and Breadth of Symptomatic Experience

Jennifer J. Newson^{1}, PhD, Dhanya Parameshwaran¹, PhD and Tara C. Thiagarajan¹, PhD.*

¹Sapien Labs, 1201 Wilson Blvd, 27th floor, Arlington, Virginia, United States of America

*** Corresponding Author**

Jennifer J Newson
Sapien Labs
1201 Wilson Blvd, 27th floor
Arlington, Virginia
USA
Phone: 1 7039976657
Email: jennifer@sapienlabs.org

ABSTRACT

Background:

The Patient Health Questionnaire-9 (PHQ-9) is the most widely used depression screener but debate still remains around its relationship to symptom experience and life impact.

Methods:

Using a sample of 5,750 respondents, we compared symptom frequency measured by PHQ-9, to ratings of life impact and the breadth of symptomatic experience measured by the Mental Health Quotient (MHQ), a transdiagnostic assessment of mental health that captures symptoms across 10 mental health disorders on a 9-point life impact scale and includes items that map to the PHQ-9.

Results:

Frequency and life impact of individual symptoms were significantly correlated. However, there was substantial variability of life impact within each frequency category, and mean life impact for the same frequency varied significantly across symptoms. In the aggregate, only 47.7% of those with PHQ-9 “severe” scores (sum score ≥ 20) met the most severe life impact threshold while 53.3% had average life impact ratings that aligned with a value corresponding to ‘OK’ or one point worse than ‘OK’ on the 9-point scale. Finally, individuals with “severe” depression measured by the PHQ-9 had numerous additional symptoms of high life impact, as well as highly heterogeneous life impact profiles across symptoms both assessed and not assessed by the PHQ-9 (Coefficient of variation of 119% and 122% for differences across individuals).

Conclusions:

The findings demonstrate that “severe” depression as determined by PHQ-9 frequency ratings is highly permissive in the context of functional life impact which may be a more meaningful metric for the individual. In addition, these individuals experience many and varied symptoms of equivalent or greater life impact beyond those captured by the PHQ-9 and have highly heterogeneous symptom profiles both within and beyond PHQ-9 symptoms. Consequently, those with “severe” depression as assessed by the PHQ-9 cannot be considered a symptomatically consistent or homogeneous group. Thus, assessments which capture a broader, transdiagnostic range of symptoms and their life impact could aid in more precise symptom profiling and treatment pathways.

Keywords:

PHQ-9, depression, major depressive disorder, assessment, screener, mental health, transdiagnostic, MHQ.

1. BACKGROUND

The Patient Health Questionnaire-9 (PHQ-9) is one of the most widely used instruments for depression screening [1-3] across many different geographical populations and clinical contexts [4-8]. It consists of nine items that match criteria for major depression as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [9] and respondents rate each item according to how frequently they have been bothered by the symptom over the last 2 weeks.

However, despite its global uptake as a depression screener [3], there is still discussion around appropriate cut-off scores [10, 11], researcher bias [12, 13], alignment with people's symptomatic experience [14, 15], and whether it is more appropriate as a general measure of distress [16]. For example, although PHQ-9 sum scores (the sum of ratings on all items) of ≥ 10 have been shown to have a sensitivity of 88% and a specificity of 88% for major depression as assessed by clinician-administered PRIME-MD [2, 17], other studies have suggested that this cut-off of ≥ 10 may result in a greater number of false positives and overestimates of depression prevalence compared to other methods [11, 18-20], a concern also raised by the Canadian Task Force on Preventive Health Care [21]. In this context, although depression scales such as the PHQ-9 rely heavily on symptom frequency as an outcome measure [22], it is currently unclear how reported frequency values (e.g. "nearly every day") in the context of depression, relate to life impact – that is the impact or consequence that a symptom is having on someone's ability to function in everyday life. Furthermore, while the PHQ-9 is designed to screen specifically for depression, other studies have suggested that the assessment may miss symptoms that are meaningful to patients [14], while, more generally, many patients typically experience a heterogeneous array of symptoms that span multiple DSM-5 defined disorders [23-25].

In this study, we therefore aimed to determine how PHQ-9 outcome measures, based on symptom frequency, related to life impact and to understand the heterogeneity of the broader symptomatic experience. For comparison we used the Mental Health Quotient (MHQ), a transdiagnostic assessment of mental health that assesses 47 individual symptoms and aspects of mental functioning, including items that map to PHQ-9 items, and provides a comprehensive view of symptomatic experience across 10 common DSM-5 defined disorders [22, 26, 27]. It also uses a 9-point life impact scale that reflects the consequences and impact on ones' life functioning rather than symptom frequency or severity. The MHQ was developed to overcome some of the limitations with existing measures, which typically aren't tailored for the general population, don't include positive assets, don't cover the full breadth of mental health symptoms and are highly heterogeneous in the way they ask about symptoms from frequency to severity and duration [22]. By comparing responses to the PHQ-9 against responses to the MHQ in the same population sample, we examined (i) the degree of equivalence between PHQ-9 frequency ratings and MHQ life impact ratings at the level of individual questionnaire items, (ii) how the PHQ-9 definition of "severe" depression (equivalent to PHQ-9 sum scores ≥ 20) related to equivalent life impact outcomes and (iii) the broader symptom profiles of individuals with "severe" PHQ-9 sum scores, to determine the scope and homogeneity/heterogeneity of their symptomatic experience.

2. METHODS

2.1. Data Acquisition

We utilized data acquired as part of the ongoing Global Mind Project (previously the Mental Health Million project) that aims to track evolving mental health and wellbeing on a global scale and currently spans 65 countries and 9 languages [28]. Participants were recruited via Ad campaigns on Facebook and Google, that targeted each adult age and gender group and directed them to the MHQ website [29]. This online recruitment method provided a rapid, flexible, low-cost and anonymized way of sampling a broad cross section of the general population aged 18+ [30]. Respondents took the assessment for the purpose of obtaining their personalized mental health report and no financial compensation was provided. The personal report aimed to ensure greater interest of the respondent in completing the 15-minute assessment and answering questions thoughtfully and accurately.

For this study, we utilized Global Mind data from 6,005 respondents collected between August and September 2022. All respondents completed the English version of the MHQ assessment as well as the PHQ-9 within the same online survey. Only respondents who responded “Yes” to the MHQ question “Did you find this assessment easy to understand?” and completed the survey in a time frame appropriate for reading all questions (≥ 7 minutes) were included in the analysis, leading to a final sample size of 5,750. Respondents were predominantly from 16 countries (see Supplementary Table 1) with the greatest proportion from United States (16.8%) and United Kingdom (10.3%). Across all countries, 42.6% and 56.6% of respondents reported their biological sex as male and female, respectively. The sample covered all age brackets with 62.1% aged 45-74 and 29.4% aged 18-44. The older skew likely reflected the self-selected nature of the sample.

2.2. The MHQ and its 47 rated items

The MHQ is a transdiagnostic assessment that comprehensively covers symptoms across 10 major mental health disorders as well as items derived from Research Domain Criteria (RDoC) [22, 26]. The list of MHQ items was determined

based on a comprehensive coding of mental health symptoms assessed in questions across 126 different mental health questionnaires and interviews (see [22] for more details and a full list of the 126 assessment tools). These included questionnaires for depression, anxiety, bipolar disorder, attention-deficit/hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), addiction, psychosis, eating disorder, and autism spectrum disorder (ASD). These disorders were selected based on their inclusion in the DSM-5 clinical interview (SCID-CV) [31]. In addition, ASD and eating disorder were included due to both their prevalence and their broad public and scientific interest. A total of 10,154 questions were coded based on semantic content and consolidated into a set of 43 symptom categories, described in more detail in [22]. The resultant items were then reviewed in the context of other transdiagnostic frameworks including RDoC constructs and subconstructs put forward by the NIMH [32, 33] and a few additions (e.g., selective attention, coordination) were made to ensure that the list of items reflected components within this non-DSM framework. The resulting categories were then reorganized into a set of 47 elements that comprehensively describe mental health and mental well-being.

Within the MHQ, each of these 47 items were rated by respondents using a 1-9 life impact scale (i.e. a Likert scale with 9 positions) reflecting the impact on one’s ability to function [26]. For items on a spectrum from positive to negative (spectrum items such as memory) 1 on the 9-point scale referred to “Is a real challenge and impacts my ability to function”, 9 referred to “It is a real asset to my life and my performance” and 5 referred to “Sometimes I wish it was better, but it’s ok”. For items with varying degrees of problem severity (problem items such as suicidal thoughts) 1 on the 9-point scale referred to “Never causes me any problems”, 9 referred to “Has a constant and severe impact on my ability to function” and 5 referred to “Sometimes causes me difficulties or distress but I can manage”. Respondents made rating responses based on their current perception of themselves.

2.3. Inclusion of PHQ-9 Questions within the Global Mind Survey

For the purpose of this study, the nine questions from the PHQ-9 were added to the same survey after MHQ scored questions such that respondents completed both assessments in sequence. Each PHQ-9 item was rated on a frequency scale of 0 to 3 that reflected how much a symptom had bothered them over the last 2 weeks (0=not at all; 1=several days; 2=more than half the days; 3=nearly every day). The sum of these ratings, the PHQ-9 sum score, was calculated for each respondent, and the proportion of respondents within each category (none=0-4; mild=5-9; moderate=10-15; moderately severe=15-19; severe \geq 20) was calculated.

2.4. Comparison between symptom frequency and life impact for individual questionnaire items

To enable comparisons of frequency ratings in the PHQ-9 and life impact ratings in the MHQ at the level of individual questionnaire items, relevant MHQ items were mapped to equivalent PHQ-9 items (Table 1). For six PHQ-9 items there was a 1-1 match with an equivalent MHQ item (“trouble falling or staying asleep, or sleeping too much/“sleep quality”; “feeling tired or having little energy”/“energy level”; “poor appetite or overeating”/“appetite regulation”; “trouble concentrating on things, such as reading the newspaper or watching television”/“focus and concentration”; “feeling down, depressed, or hopeless”/“feelings of sadness, distress or hopelessness”; “thoughts that you would be better off dead, or of hurting yourself in some way”/“suicidal thoughts or intentions”). To determine how symptom frequency related to life impact for each 1-1 match, we calculated the Pearson correlations between the PHQ-9 rating and MHQ rating, as well as the average life impact rating of each MHQ item for each symptom frequency response category (e.g., “nearly every day”) from the equivalent PHQ-9 item. We then used a standard t-test to determine the statistical significance of differences in the average MHQ life impact ratings of each item corresponding to each successive PHQ-9 rating selection, and between different MHQ items corresponding to PHQ-9 ratings of “nearly every day”.

Table 1:
MAPPING OF MHQ ITEMS TO PHQ-9 ITEMS

MATCH	PHQ-9 ITEM	MHQ ITEM (S) (S = SPECTRUM ITEM; P = PROBLEM ITEM)
1-2	Little interest or pleasure in doing things	Drive and motivation (S); Curiosity, interest and enthusiasm (S)
1-1	Feeling down, depressed, or hopeless	Feelings of sadness, distress or hopelessness (P)
1-1	Trouble falling or staying asleep, or sleeping too much	Sleep quality (S)
1-1	Feeling tired or having little energy	Energy level (S)
1-2	Feeling bad about yourself or that you are a failure or have let yourself or your family down?	Self-worth and confidence (S); Guilt and blame (P)
1-1	Poor appetite or overeating	Appetite regulation (S)
1-1	Trouble concentrating on things, such as reading the newspaper or watching television	Focus and concentration (S)
1-2	Moving or speaking so slowly that other people could have noticed. Or the opposite being so fidgety or restless that you have been moving around a lot more than usual?	Confusion or slowed thinking (P); Restlessness and hyperactivity (P)
1-1	Thoughts that you would be better off dead, or of hurting yourself in some way	Suicidal thoughts or intentions (P)

2.5. Life impact profiles of those with “severe” depression as defined by the PHQ-9

Within the PHQ-9, symptom frequency ratings from the 9 items are summed to create an aggregate score that is categorized from “none” to “severe” depending on the score window (see above). To determine how “severe” depression, based on an aggregate of symptom frequency (PHQ-9 sum scores of ≥ 20), related to “severe” depression based on an aggregate of life impact, we first constructed a metric (hereafter called the MHQ-depression-9) that aligned with the complete PHQ-9. This metric was composed of the six 1-1 matched MHQ items described above, as well as a subset of MHQ item pairs that aligned with the remaining 3 PHQ-9 items. For example, the MHQ items of “drive and motivation” and “curiosity, interest and enthusiasm” were matched with the PHQ-9 item of “little interest or pleasure in doing things” (see Table 1 for the complete mapping between the PHQ-9 and MHQ-depression-9, and the limitations section for a discussion of this mapping procedure). We then transformed the MHQ 1 to 9 life impact scale into a 0 to 3 scale, to align with the scale used in the PHQ-9, using 3 different transformations that varied in terms of how life impact mapped onto symptom frequency (T1, T2, T3; see Figure 3). T1 was most stringent in the life-impact definition of “severe” (where ratings of ≥ 8 for problem items and ≤ 2 for spectrum items were mapped to “nearly every day”), T2 was intermediate, (≥ 7 for problem items and ≤ 3 for spectrum items) and T3 the most permissive (≥ 6 for problem items and ≤ 4 for spectrum items). In addition, due to the differences in 1 to 9 scales for spectrum and problem items, different transformations were used for each. For the 3 PHQ-9 items where two MHQ items were matched to one PHQ-9 item, the two life impact ratings were transformed, and the average calculated. PHQ-9 sum scores, and sum scores for this constructed MHQ-depression-9 metric, were then calculated and compared for these 3 transformation thresholds to determine how permissive or stringent the PHQ-9 definition of “severe” depression was in relation to impact on people’s ability to function in life.

2.6. Assessing heterogeneity of symptom profiles for individuals with “severe” PHQ-9 sum scores

To explore the wider symptomatic experience of individuals who would be considered at risk of depression according to their PHQ-9 sum score, we examined the broader symptom profiles, collected using the MHQ, of individuals with “severe” PHQ-9 sum scores (scores ≥ 20 ; N=373), including the 35 items assessed by the MHQ that do not map to PHQ-9 items. To determine the prevalence of symptoms for each of the 35 items not mapped to the PHQ-9 and the 12 MHQ items that mapped to the PHQ-9, we calculated the percentage of respondents who selected ratings of >7 for problem items or >3 for spectrum items, denoting the two most serious negative life impact ratings. To estimate the degree of difference in the symptom profile between each pair of individuals with “severe” PHQ-9 score, we calculated the absolute difference in life impact ratings for each pair of respondents as shown below:

MHQ ITEM	1	2	3	4	5	6	7	8	Etc.
Life impact rating (respondent 1)	1	5	7	9	5	6	7	3	Etc.
Life impact rating (respondent 2)	3	7	8	3	5	5	4	8	Etc.
Absolute Difference	2	2	1	6	0	1	3	5	Etc.

As measures of heterogeneity, we then computed the mean and standard deviation (SD) of these absolute differences for each pair of individuals for the 35 MHQ items not matched to the PHQ-9, as well as for the group of 12 MHQ items that were matched to PHQ-9 items. We then looked at whether the distributions of the means and SDs for the PHQ-9 matched and non-matched items were different by computing statistical significance with a two-sample t-test.

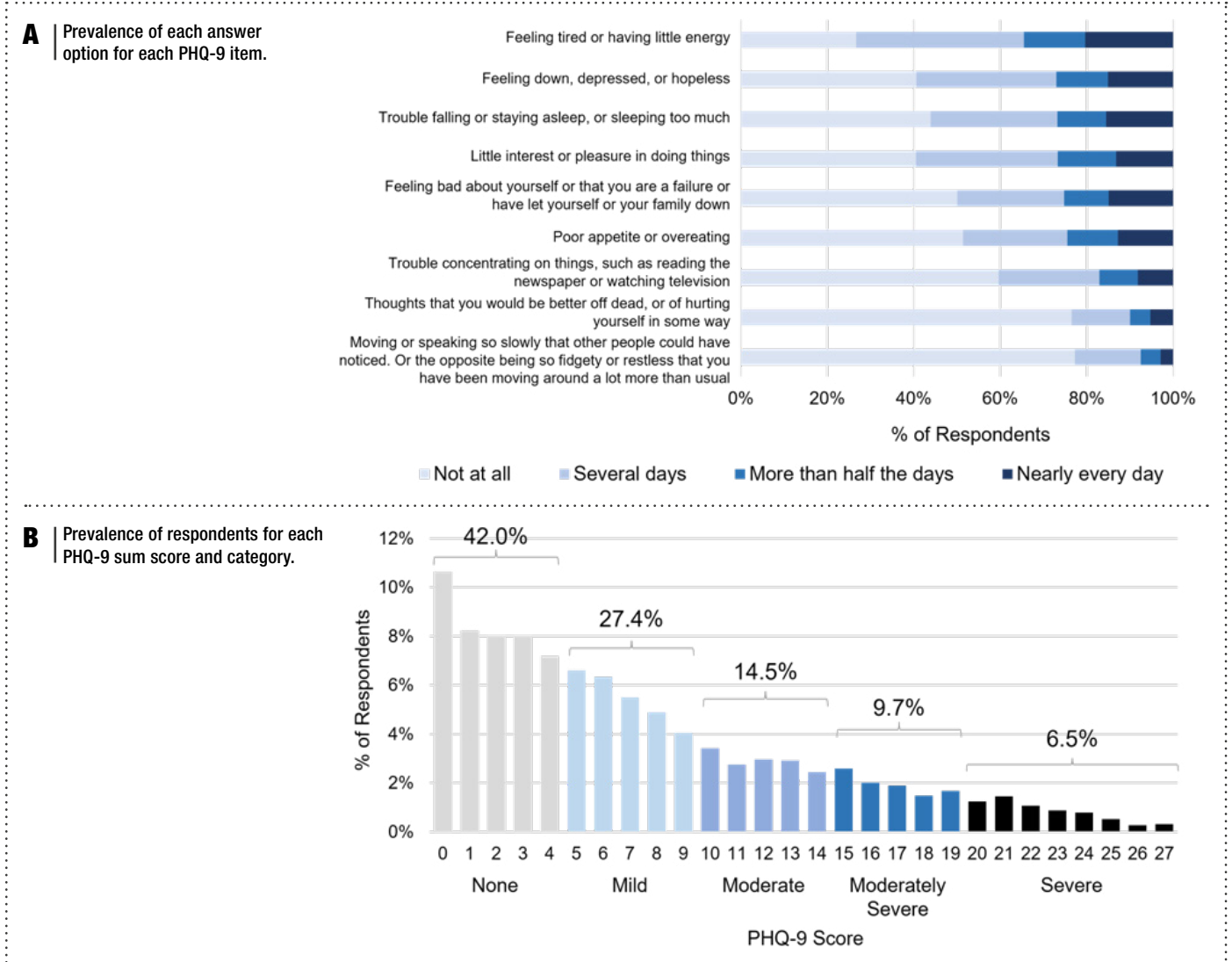
3. RESULTS

3.1. Prevalence and distribution of PHQ-9 answers and scores

To start with, we looked at the frequency of answer selections for each PHQ-9 item or symptom within this general population sample to determine their prevalence, as well as the distribution of PHQ-9 sum scores. Figure 1A shows the prevalence of each frequency rating for each PHQ-9 item within the sample. The item “feeling tired or having little energy” had the greatest proportion of respondents with ratings of “more than half the days” (14.1%) and “nearly every day” (20.4%), while the item of “moving or speaking so slowly that other people could have noticed. Or the opposite, being so fidgeting or restless that you have been moving around a lot” had the lowest proportion of respondents with ratings

of “more than half the days” (4.6%) and “nearly every day” (2.9%), which is broadly in line with other findings [34]. Thus, there was considerable variability in the prevalence of individual symptoms. The distribution of PHQ-9 sum scores based on these frequency ratings are shown in Figure 1B. 42.0% of respondents had summed PHQ-9 scores of 0-4 (PHQ-9 category “none”), while 6.5% had sum scores ≥ 20 (PHQ-9 category “severe”) with the most prevalent symptoms in this “severe” category (i.e. highest percentage of “nearly every day” responses) being “feeling tired or having little energy” (90.1%), “feeling down, depressed, or hopeless” (87.4%) and “feeling bad about yourself or that you are a failure or have let yourself or your family down” (84.5%). Overall, 30.6% had PHQ-9 sum scores ≥ 10 , considered to indicate mild through to severe depression [2].

Figure 1 - DISTRIBUTION OF PHQ-9 ANSWERS AND SCORES WITHIN THE SAMPLE



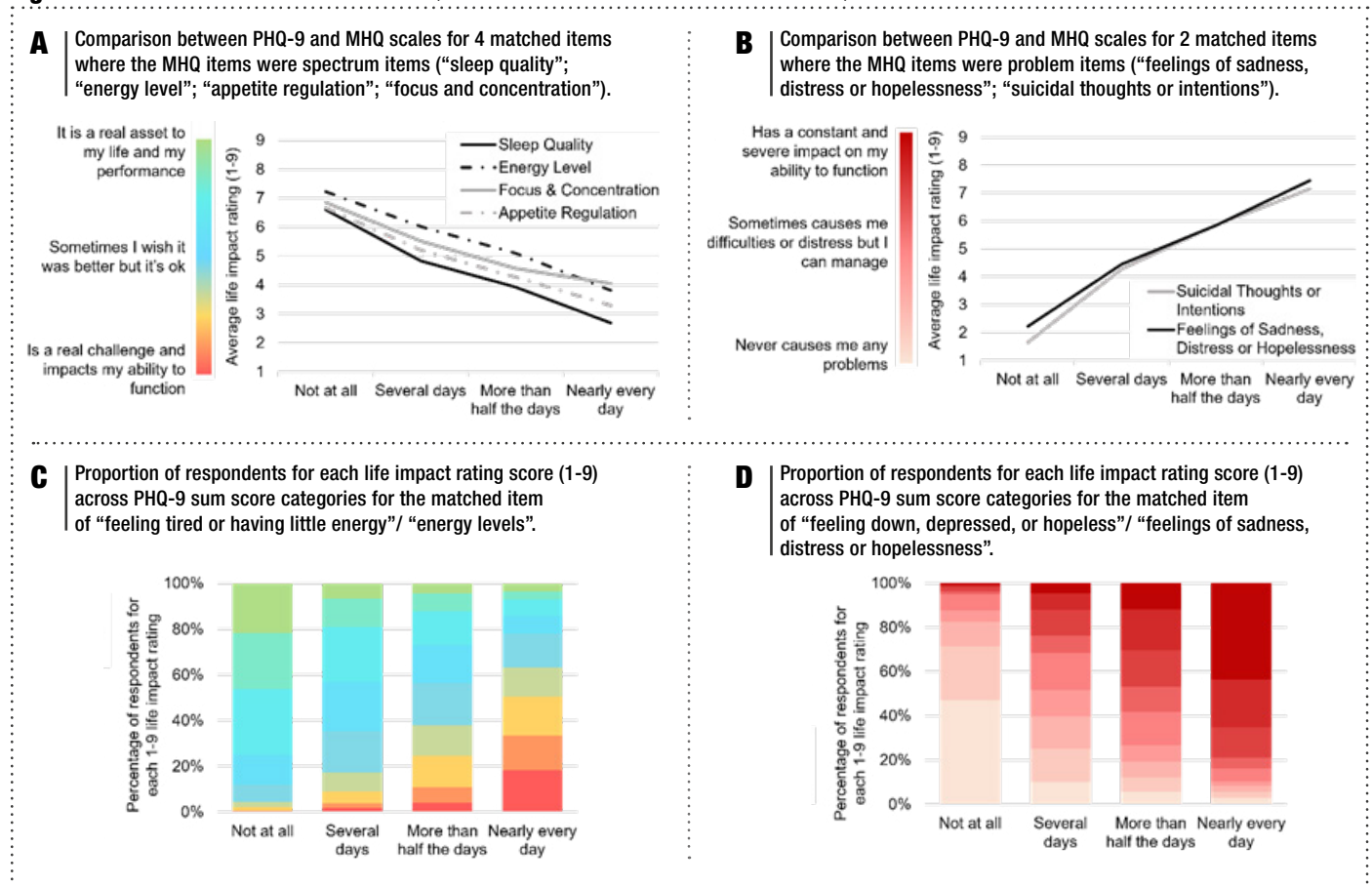
3.2. Comparison between symptom frequency and life impact for individual questionnaire items

The relationship between symptom frequency and life impact was examined for the 6 items where there was a 1-to-1 match between the PHQ-9 and MHQ (Figure 2; also see methods and Table 1). For each item, the average MHQ life impact rating (1-9) was calculated for each PHQ-9 frequency rating category from “not at all” to “nearly every day”. For example, the average life impact rating for the MHQ item “energy levels” was calculated for those respondents who answered “nearly every day” on the equivalent PHQ-9 item of “feeling tired or having little energy as not at all”. For the four PHQ-9 items that were mapped to MHQ spectrum items (where 1 was “A real challenge and impacts my ability to function” and 9 was “It is a real asset to my life and my performance”) a PHQ-9 rating of “nearly every day” was associated with

average life impact ratings that ranged from 2.7 ± 1.9 (average \pm SD) for “trouble falling or staying asleep, or sleeping too much”/“sleep quality” to 4.0 ± 2.3 for “trouble concentrating on things, such as reading the newspaper or watching television”/“focus and concentration”, with an overall average across all 4 items of 3.5 ± 2.1 , and were statistically different from each other ($p < 0.001$) for all pairs except Focus & Concentration and Energy Levels (Supplementary Table 2). In contrast, the PHQ-9 rating of “not at all” was associated with average life impact ratings that ranged from 6.6 ± 0.3 for “trouble falling or staying asleep, or sleeping too much”/“sleep quality” to 7.2 ± 0.3 for “feeling tired or having little energy”/“energy level” with an overall average across all items of 6.8 ± 0.3 .

For the 2 PHQ items that were mapped only to MHQ problem items (where 1 was “Never causes me any problems”, and 9 was “Has a constant and severe

Figure 2 - COMPARISON OF MHQ LIFE IMPACT SCALE AND PHQ-9 RATING SCALE



impact on my ability to function”; Table 1), a PHQ-9 rating of “nearly every day” for “thoughts that you would be better off dead, or of hurting yourself in some way” was associated with an average life impact rating of 7.1 ± 2.1 for “suicidal thoughts or intentions” while a PHQ-9 rating of “nearly every day” for “feeling down, depressed, or hopeless” was associated with an average life impact rating of 7.4 ± 2.4 for “feelings of sadness, distress or hopelessness”, with an average across both items of 7.3 ± 2.3 . In contrast the PHQ-9 rating of “not at all” was associated with average life impact ratings of 1.6 ± 0.1 for “suicidal thoughts or intentions” and 2.2 ± 0.1 for “feelings of sadness, distress or hopelessness” (overall average 1.9 ± 0.1).

Thus, on average, negative life impact increased significantly with increasing frequency for each symptom ($p < 0.001$ between means of each successive PHQ-9 frequency category by standard t-test; Supplementary Table 3). However, there was much higher variability of life-impact ratings for the “nearly every day” PHQ-9 rating group with standard deviations of 2 points or more compared to the “not at all” group where standard deviations were less than 0.5 points. Therefore, high symptom frequency can have a wide range of life impact (i.e., the person’s ability to function in life) at an individual level that has relevance in terms of individual clinical evaluation.

Figure 2 compares the proportion of responses for each life impact ratings (1-9) for the two PHQ-9 items with greatest prevalence of severe ratings (“feeling tired or having little energy”, Figure 2C; and “feeling down, depressed, or hopeless”, Figure 2D; additional items shown in Supplementary Figure 1). Of the participants who responded to the PHQ-9 item of “feeling tired or having little energy” with the highest frequency rating of “nearly every day”, 50.6% also responded to the MHQ matched item of “energy levels” with a life impact rating of 1, 2 or 3, reflecting a negative impact, while 21.9% of these participants responded with life impact ratings >5 indicating that despite experiencing feeling tired or having little energy nearly every day, this did not

have much impact on their ability to function in life and was even an asset to their life. Conversely, of the participants who responded to this PHQ-9 item with the lowest frequency rating of “not at all”, 75.0% also responded to the MHQ matched item of “energy levels” with a life impact rating of 7, 8 or 9, reflecting it being an asset in their life and performance, while less than 1% had life impact ratings less than 3.

For the PHQ-9 item of “feeling down, depressed, or hopeless”, 78.9% of respondents who rated it as “nearly every day”, rated the MHQ matched item of “feelings of sadness, distress or hopelessness” with a life impact rating of 7, 8 or 9, reflecting a serious impact, while 82.6% of respondents who rated this same item as “not at all”, on the PHQ-9 rated the MHQ matched item as 1, 2 or 3, reflecting it rarely causing problems to function (Figure 2D). However, while these extremes separated well in terms of life impact, selections of “several days” and “more than half the days” on the PHQ-9 scale had broad ranging life impact across the MHQ scale from 1-9 where a selection of 3 on the life impact scale was as prevalent as a selection of 7.

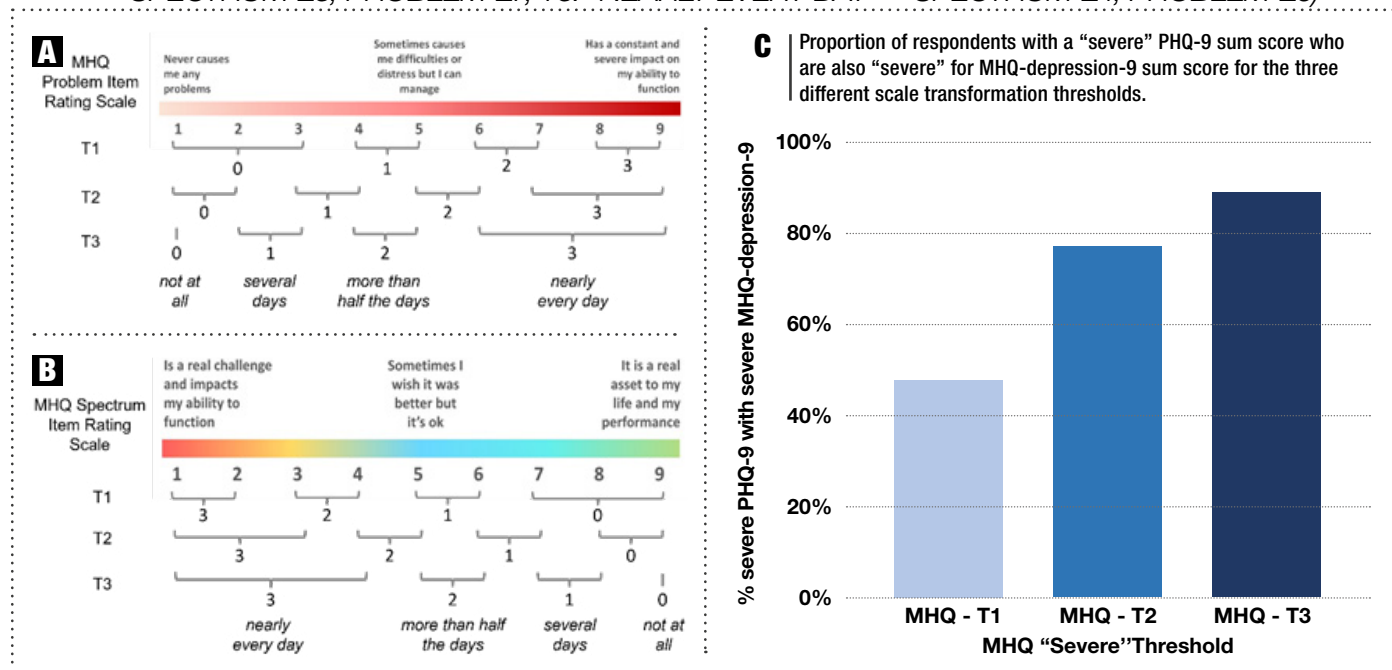
The association between PHQ-9 frequency ratings and MHQ life impact ratings for each individual item was evaluated using Pearson correlations and ranged from 0.67 for “thoughts that you would be better off dead, or of hurting yourself in some way”/“suicidal thoughts or intentions” and “feeling down, depressed, or hopeless”/“feelings of sadness, distress or hopelessness” to 0.47 ($p < 0.001$ all cases) for “trouble concentrating on things, such as reading the newspaper or watching television”/“focus and concentration”, respectively. While these correlations are statistically significant given the large N, they nonetheless convey that how much an individual was suffering in terms of life impact is not adequately captured by the frequency scale. Together with the statistical differences between mean life impact ratings of each item for the “nearly every day” category, it shows that different symptoms experienced with the same frequency have different life impact.

3.3. Life impact equivalence of “severe” PHQ-9 sum scores

The PHQ-9 draws arbitrary lines to delineate between “severe” and “moderately severe” depression which can then determine treatment pathways. Within the sum scores itself, as shown in Figure 1B, changing the threshold for “severe” by even one point can change the prevalence by ~3%. Here we look at how permissive or stringent the PHQ-9 definition of “severe” depression was in relation to impact on people’s ability to function in life. This is important at an individual level if the goal is to ensure that all patients with a severe life impact of symptoms are captured and, conversely, to ensure that those who are classified as “severe” according to sum scores of symptom frequency do indeed have a severe functional impact. Thus, to determine the life impact threshold that, in the aggregate, best captured the “severe” PHQ-9 category, we compared sum scores for those MHQ elements that mapped to the PHQ 9 (MHQ-depression-9 metric) constructed using three different rating transformations that varied in the thresholds used to map life impact to symptom frequency (Figure 3A and 3B). For example, T1 represents an MHQ-depression-9 score constructed

by mapping a life impact of >7 on the problem scale and <3 on the spectrum scale to the PHQ-9 equivalent of “nearly every day” (the most stringent mapping). In contrast T3 represents a much more permissive mapping of >5 on the problem scale (where 5 is “Sometimes causes me distress but I can manage”) and <5 on the spectrum scale (where 5 is “Sometimes I wish it was better but its OK”). We then looked at what percentage of those in the “severe” PHQ-9 category also had a “severe” MHQ-depression-9 sum score for each threshold, essentially the precision of the PHQ-9 for identifying individuals experiencing severe life impact of symptoms (Figure 3C). At the most stringent threshold (T1), precision was 47.7%, while sensitivity or recall was 58.4%. At the most permissive threshold (T3), while the precision increased to 89.0%, sensitivity decreased to 27.3% reflecting the much larger percentage classified as having “severe” life impact. On the other hand, specificity was high at 96.4% for T1 and 99% for T3. Altogether, even with “severe” PHQ-9 defined as a sum score of ≥ 20 , a third of those classified as having “severe” depression by the PHQ-9 had an average life impact rating of symptoms that was either ‘OK’ or just a bit worse than ‘OK’.

Figure 3 - TRANSFORMATION OF SYMPTOM FREQUENCY RATING TO LIFE IMPACT RATING FOR PROBLEM (A) AND SPECTRUM (B) ITEMS FOR THREE TRANSFORMATION THRESHOLDS (T1: “NEARLY EVERY DAY” = SPECTRUM ≤ 2 ; PROBLEM ≥ 8 ; T2: “NEARLY EVERY DAY” = SPECTRUM ≤ 3 ; PROBLEM ≥ 7 ; T3: “NEARLY EVERY DAY” = SPECTRUM ≤ 4 ; PROBLEM ≥ 6)

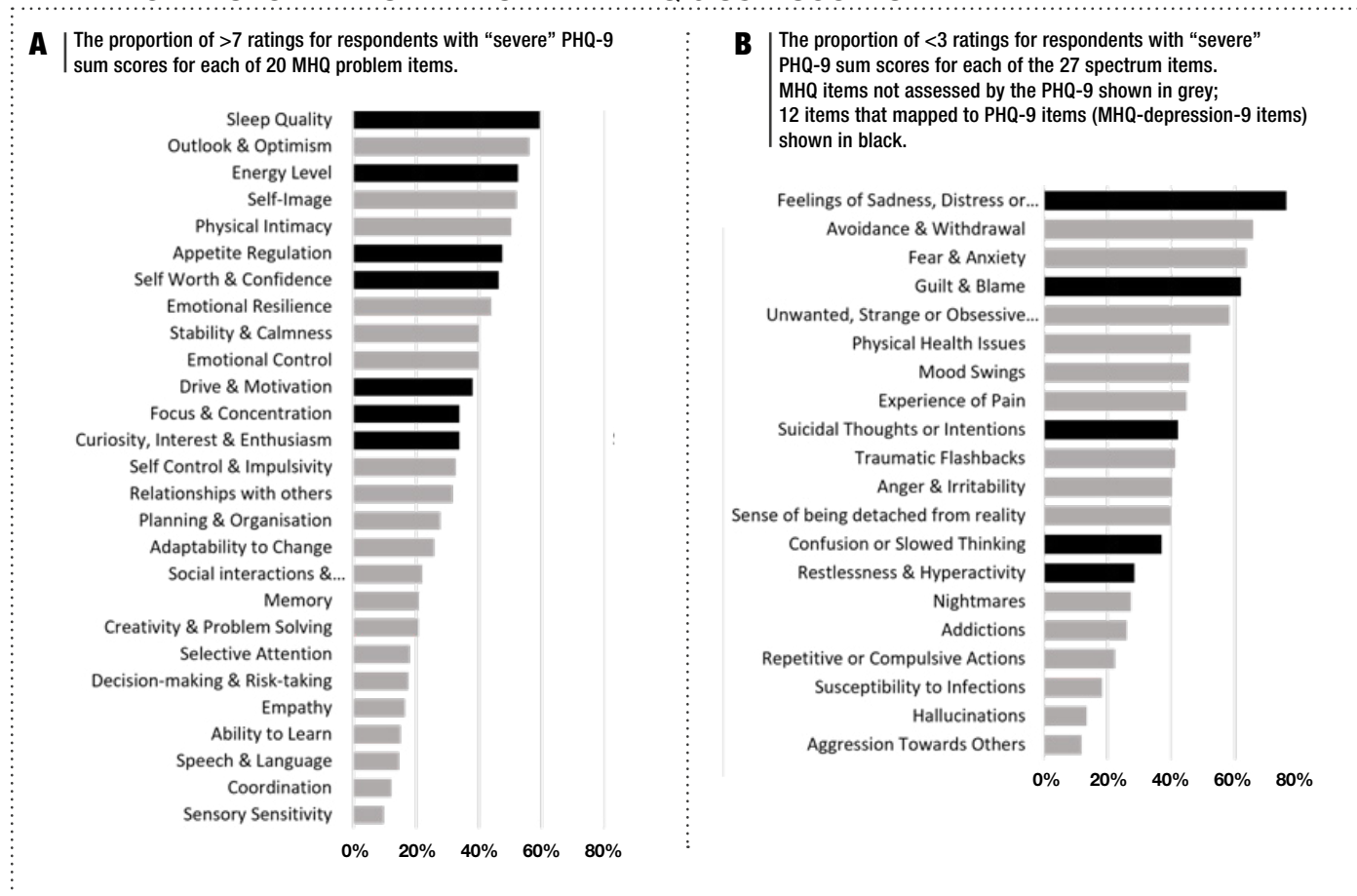


3.4. Prevalence of symptoms beyond the PHQ-9 for respondents with “severe” PHQ-9 sum scores.

To examine the wider symptomatic experience of those individuals classified as having “severe” depression by the PHQ-9, we next compared the full symptom profiles of only those respondents with “severe” PHQ-9 sum scores. Figure 4 shows the proportion of respondents within this subgroup who rated each MHQ problem item >7 (Figure 4A) or spectrum item <3 (Figure 4B; i.e. denoting that they are having a serious impact on someone’s ability to function). The 12 items that mapped to PHQ-9 items are shown in black (MHQ-depression-9 items) while the 35 MHQ items not assessed by the PHQ-9 are shown in grey. Within the spectrum items, several items not assessed by the PHQ-9, including “outlook and optimism” (56.0%), “self-image” (52.0%) and

“physical intimacy” (50.4%) ranked at similarly high prevalence levels to MHQ-depression-9 items (e.g., “sleep quality”, 59.5%; “energy level”, 52.5%; “appetite regulation”, 47.5%). A similar pattern emerged for problem items, where “avoidance & withdrawal” (65.7%), “fear & anxiety” (63.5%), “unwanted, strange or obsessive thoughts” (58.2%) ranked at similarly high levels of prevalence as MHQ-depression-9 items of “feelings of sadness, distress or hopelessness” (76.1%) and “guilt & blame” (61.9%). This suggests that there were several commonly occurring symptoms having a serious impact on people’s ability to function in this “severe” PHQ-9 subgroup, that were not assessed by the PHQ-9. The high prevalence of symptoms relating to anxiety is also in line with other reports suggesting high levels of comorbidity between depression and anxiety [24].

Figure 4 - PREVALENCE OF SEVERE NEGATIVE LIFE IMPACT RATINGS ACROSS 47 MHQ ITEMS FOR RESPONDENTS WITH “SEVERE” PHQ-9 SUM SCORES

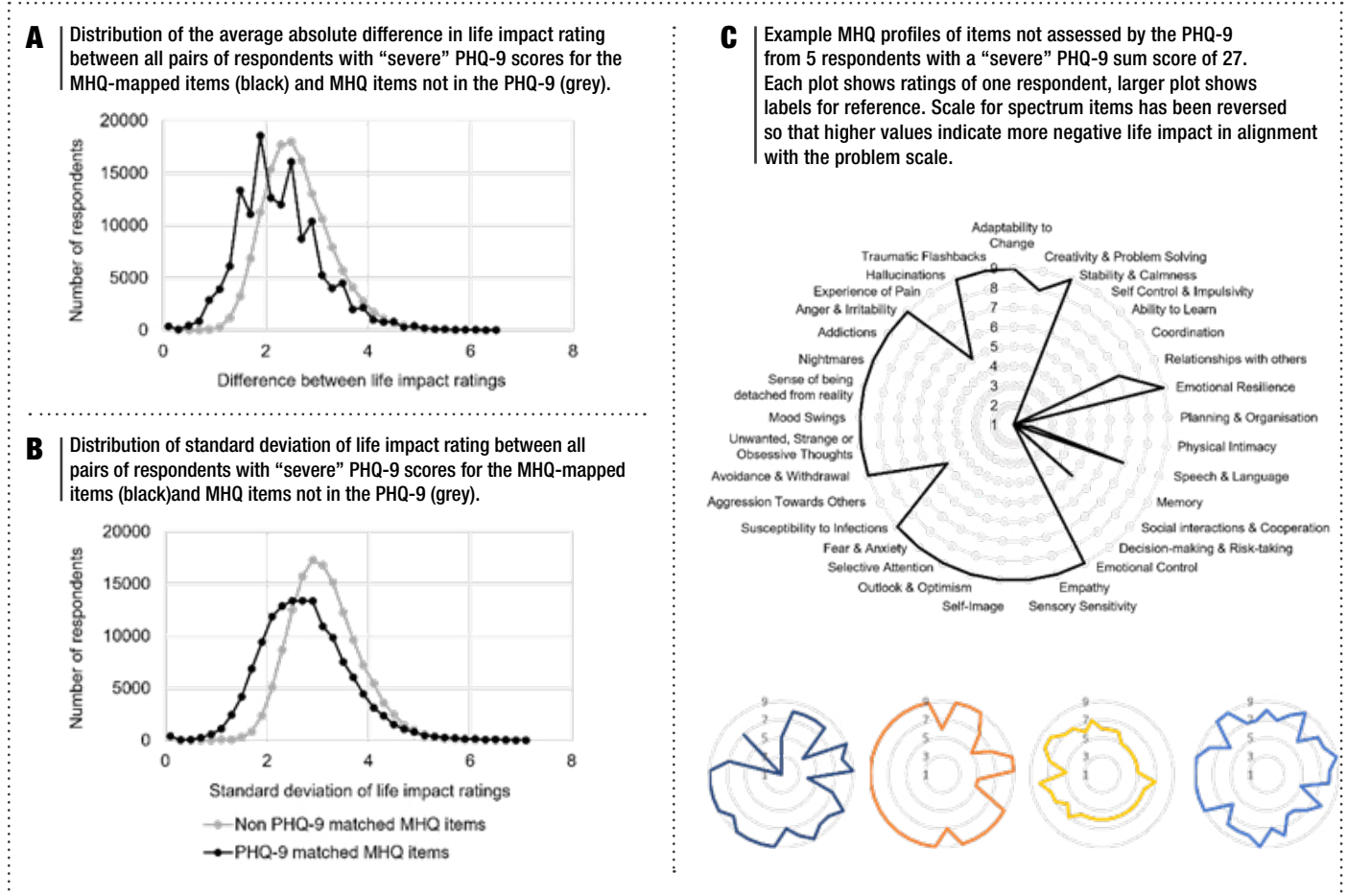


3.5. Heterogeneity of symptom profiles

We next looked at how heterogeneous the symptom profiles were for those within the “severe” PHQ-9 group. We did this by examining the differences in life impact ratings for the group of 12 MHQ items that mapped to the PHQ-9 (MHQ-depression-9 items) as well as the group of additional 35 MHQ items. Figure 5A shows the distribution of the average absolute difference in life impact rating between all pairs of respondents with “severe” PHQ-9 scores for MHQ items not mapped to the PHQ-9 (grey) as well as for MHQ-depression 9 items (black) while Figure 5B shows the corresponding distributions of standard deviations in the differences between life impact ratings. For MHQ items not mapped to the PHQ-9 the average difference in life impact ratings was 2.6 ± 3.1 . In comparison, for MHQ-depression-9 items the average difference in life impact ratings was 2.3 ± 2.8 . While the life impact differences between

these two groups were statistically significant ($p < 0.001$ by t-test) indicating that PHQ-9 matched items had lower mean differences in life impact than those items not within the PHQ-9, there was considerable variability within both groups. Coefficient of variation (CV) was 122% in percentage terms for PHQ-9 matched items and 119% for the non-matched items. Ratings varied as much as 5 to 6 points on the 9-point life impact scale for individual items. As an illustrative example of this heterogeneity, Figure 5C shows the MHQ symptom profiles of the 35 items not assessed by the PHQ-9 from 5 respondents, selected at random from the group of 33 respondents with the maximum possible PHQ-9 sum score of 27. Overall, this suggests that individuals within the “severe” PHQ-9 group show considerable variability in their broader symptom profile, and commonly experience functionally impacting symptoms that are not assessed by the PHQ-9.

Figure 5 - VARIABILITY OF MHQ PROFILES FOR RESPONDENTS WITH “SEVERE” PHQ-9 SUM SCORES



4. CONCLUSIONS

This study describes how the self-reported frequency of depression symptoms, as measured by the PHQ-9, relates to ratings of life impact for equivalent items, as measured by the MHQ, and characterizes the breadth of symptom experience of individuals with “severe” PHQ-9 scores, capitalising on the comprehensive, transdiagnostic nature of the MHQ assessment. Key results are that “severe” depression, as measured by the PHQ-9 sum score, is highly permissive in its life impact, and that symptomatic experience is highly heterogeneous within both the depression-specific symptoms assessed by the PHQ-9 and the 35 additional symptoms captured by the MHQ.

4.1. Symptom frequency is not consistently equivalent to life impact

Although on average, increasing frequency of symptoms was associated with increasing life impact, at the level of individual items the relationship between symptom frequency and life impact was inconsistent. For example, a PHQ-9 rating of “nearly every day” for “feeling down, depressed, or hopeless” had a high average life impact rating of 7.2 but a considerable range from 1 to 9 among individuals. This could arise due to varied interpretation of “nearly every day” as anywhere from 4-7 days per week as well as differences in the severity of impact, which is also correlated with life impact [27]. For example, one person may experience mild symptoms frequently and rate it as “nearly every day” on the PHQ-9 but as low on a life impact scale. Conversely someone may have symptoms infrequently but experience them severely and therefore rate it as “some of the time” on the PHQ-9 but high in its life impact. Such variability was most prominent for the “several days” and “more than half the days” selections. Furthermore, there was considerable variability across the life impact of different PHQ-9 matched items. For example, average life impact for “energy levels” was lower compared to “sleep quality” for the equivalent PHQ-9 ratings. This suggests that the level of functional impact is not equivalent across symptoms and that simple sums of symptom frequency may not appropriately capture true life experience or consequence.

Altogether, this variability speaks to the wider consideration of which symptom criteria are most important from a clinical decision-making perspective, and how different rating scales can lead to different

outcomes. This is important given the heterogeneity of mental health assessments which differentially capture various aspects from frequency to severity and duration, all of which contribute to life impact [22, 35].

4.2. The importance of thresholds

There has been considerable discussion in the literature around the appropriate PHQ-9 sum score threshold for clinical relevance with some suggesting that a PHQ-9 cut off threshold of ≥ 10 may overestimate depression prevalence [10, 11, 20]. Such debates also contribute to wider discussions around when symptoms should be considered clinically significant versus being part of the natural ups and down of life [36, 37], particularly in the context of the stark rise in antidepressant prescribing [38]. In this sample from a broad general population, PHQ-9 sum scores followed a long tail distribution where a threshold of ≥ 10 meant that 30% of the sample would be considered at risk for depression. However, as would be expected, raising the transformation threshold by even 1-point reduced prevalence estimates of depression by 3-4%. While a PHQ-9 cut off threshold of ≥ 10 is typically considered to offer the best balance between sensitivity and specificity in some contexts (e.g. primary care) [2, 19] these findings reinforce how small shifts in this threshold have considerable implications for prevalence estimates, clinical decision-making and the risk of false positives/ negatives within a clinical setting [10].

The results here show that the PHQ-9 definition of “severe” depression using even the higher sum score of ≥ 20 is highly permissive in terms of its equivalence to life impact. Only 47.7% of respondents within the “severe” PHQ-9 category were also “severe” on the MHQ-depression-9 at the most stringent life impact threshold, while 89.0% of respondents were “severe” on the MHQ-depression-9 at the least stringent life impact threshold. Thus, almost a third of those with “severe” depression, as determined by a PHQ-9 sum score of ≥ 20 , had an average life impact of symptoms that was either ‘OK’ or just a single point worse than ‘OK’ on the 9-point life-impact scale. Conversely, 29.4% with severe life impact of PHQ-9 symptoms were not classified as “severe” by the PHQ-9. We suggest that the life-impact of symptoms may be a more meaningful outcome to track in individuals during clinical evaluation and treatment.

4.3. “Severe” PHQ-9 sum scores encompass diverse symptomatic experience

Although the PHQ-9 is a depression screener, there is also debate as to whether it misses some meaningful symptoms [14] and how it related to wider symptomatic or comorbid experiences. Within this sample, 5 out of 9 PHQ-9 items were experienced by the majority of respondents on several days or more out of the last 2 weeks suggesting these symptoms were fairly ubiquitous in the population. However, within the subgroup of respondents in the “severe” depression category, several symptoms not assessed by the PHQ-9 had an even higher prevalence or greater impact on life function than many of the PHQ-9 symptoms. For example, 63-66% of respondents with “severe” depression reported highly problematic life impact ratings for anxiety-related symptoms of “avoidance & withdrawal” and “fear & anxiety”, while, in contrast, only 28% had the PHQ-9 matched symptom of “restlessness & hyperactivity”. Altogether, these results are in line with findings of high levels of comorbidity between depression and anxiety [24, 39, 40] and add to previous evidence suggesting that the PHQ-9 may miss the presence or intensity of some symptoms that are meaningful to patients [14, 15].

Some studies have interpreted this general heterogeneity of symptoms in terms of different manifestations of depression itself [41, 42]. However, another interpretation might be considered by looking at the heterogeneity of symptom profiles rather than diagnostic comorbidities. Here we have shown that the heterogeneity of life impact ratings of symptoms associated with “severe” PHQ-9 sum scores was high, with a Coefficient of Variation (CV) of 122% for the PHQ-9 matched symptoms and 115% for the 35 symptoms not captured by the PHQ-9. For those MHQ items not assessed by the PHQ-9, there was an average difference of 2.6 points between life impact ratings and an average standard deviation of 3.1 reflecting a possible range of 1 to 7, representing almost the entire breadth of the scale. This high level of heterogeneity also aligns with previous analysis showing that individuals with symptoms aligned to criteria of severe depression have overall symptom profiles that are almost as heterogeneous as symptom profiles between individuals with each of two diagnoses such as depression and ADHD [23]. This suggests that the PHQ-9 may be more relevant as a rapid screener of

general distress rather than a specific physiologically bounded construct.

Thus, while the PHQ-9 has high sensitivity and specificity for the criteria as laid out by the DSM-5 [1-3], it speaks to the growing literature that highlights the misalignment between disorder classifications and symptomatic experience, where symptoms vary considerably between individuals with the same diagnosis, shift over the lifespan and often transverse diagnostic criteria spanning multiple disorders [23, 41-46]. In turn, this has both clinical and research implications, where patients may have to embark on a long struggle to find effective treatment [47] and where studies developing new therapies and medications use diagnostic groups that may be substantially heterogeneous in terms of their symptom profiles, and therefore outcomes (e.g. [48]).

4.4. Limitations and Future Directions

Within this study, several limitations should be noted. First, the sample used in this study are self-selected respondents from a general population who are capable and interested in participating in an Internet-based self-assessment. It is therefore not a perfectly representative sample. As such, prevalence numbers at any threshold cannot be assumed to reflect population prevalence. Second, the mapping between the PHQ-9 and the MHQ was based on a best fit approach, where some PHQ-9 items aligned with two MHQ items, rather than a single item and where the MHQ assessment was based on an individual’s current perception, rather than the past 2 weeks. For example, although the PHQ-9 item of “Little interest or pleasure in doing things” mapped to two similar, broadly defined MHQ-9 items (“drive and motivation”; “curiosity, interest and enthusiasm”), this may not be perfectly equivalent. There may therefore be a degree of error in item mapping between the PHQ-9 and MHQ. This also puts a spotlight on the general challenge of semantics where word choices, although subtle can potentially have significant impact on interpretation. Third, in both cases of PHQ-9 and MHQ responses, this data reflects self-report and is therefore perceptual rather than empirically quantified. However, this is a challenge of the field of mental health in general where frequency and impact of feelings and perceptions are not empirically quantifiable. Fourth, the sample reflected a broad cross section of the general

global population, spanning multiple age groups and cultures. It is possible that the relationship between PHQ-9 outcomes based on symptom frequency and life impact/symptomatic experience may differ between age and cultural groups. It will be important to explore this in future studies, especially in light of the differences in overall mental health and wellbeing across age groups and the increased prevalence of mental health challenges in younger age groups. In addition, future research could also further explore comparisons between PHQ-9 and overall MHQ scores which provide an overall measure of mental health and wellbeing [26, 27] and could add further to the discussion around whether the PHQ-9 is better suited as an overall measure of distress [16].

Nonetheless, it is unlikely that the overall outcome of this study, that “severe” depression as determined by the PHQ-9 encompasses highly heterogeneous life impact and symptom profiles, would be changed by these limitations. Therefore, while the PHQ-9 can serve as a first level screener, subsequent assessments with transdiagnostic tools that capture a broader range of symptoms and more comprehensively reflect the comorbid nature of mental health disorders and their life impact, is an important second step to obtain a more complete picture of the individual’s symptomatic experience to aid clinical decision-making.

5. LIST OF ABBREVIATIONS

MHQ: Mental Health Quotient

PHQ-9: Patient Health Questionnaire -9

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

CV: Coefficient of variation

ADHD: attention-deficit/hyperactivity disorder

ASD: autism spectrum disorder

OCD: obsessive compulsive disorder

PTSD: post-traumatic stress disorder

RDoC: Research Domain Criteria

6. DECLARATIONS

Ethics approval and consent to participate

Research carried out on human subjects was conducted in compliance with the Helsinki Declaration and was approved by an appropriate institutional review board. Specifically, the Mental Health Million Project is a public interest project that has ethics approval from the Health Media Lab Institutional Review Board (HML IRB), an independent IRB that provides assurance for the protection of human subjects in international social and behavioral research (OHRP Institutional Review Board #00001211, Federal Wide Assurance #00001102, IORG #0000850). Participants participated in the online survey voluntarily, anonymously, and without any compensation. Participants provided electronic informed consent, were allowed to complete the questionnaire in their own time and could stop completing the survey at any time if they wanted.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in the supplementary information files associated with this article. The full dataset from the Mental Health Million Project is freely available for not-for profit purposes from the Sapien Labs Researcher Hub. Access can be requested here: <https://sapienlabs.org/mental-health-million-project/researcher-hub/>.

Competing interests

One author (TT) received a grant award from the National Institute of Mental Health (NIMH) to develop a commercial version of the MHQ tool referenced herein for clinical use. All other authors do not have any competing interest.

Funding

This work was supported by funding from Sapien Labs. It did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

All authors contributed to the design and conduct of the analyses. JN and TT participated in the interpretation of the findings and writing of the manuscript. All authors approved the final version.

Acknowledgements

With thanks to the Sapien Labs team for assistance with data infrastructure and outreach recruitment.

REFERENCES

1. Spitzer RL, Kroenke K, Williams JB: **Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire.** *Jama* 1999, 282(18):1737-1744.
2. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure.** *J Gen Intern Med* 2001, 16(9):606-613.
3. Kroenke K: **PHQ-9: global uptake of a depression scale.** *World Psychiatry* 2021, 20(1):135-136.
4. Wisting L, Johnson SU, Bulik CM, Andreassen OA, Rø Ø, Bang L: **Psychometric properties of the Norwegian version of the Patient Health Questionnaire-9 (PHQ-9) in a large female sample of adults with and without eating disorders.** *BMC Psychiatry* 2021, 21(1):6.
5. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9.** *BMC Psychiatry* 2008, 8(1):46.
6. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, Rubinstein A, Araya R, Irazola V: **Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina.** *BMC Psychiatry* 2019, 19(1):291.
7. Degefa M, Dubale B, Bayouh F, Ayele B, Zewde Y: **Validation of the PHQ-9 depression scale in Ethiopian cancer patients attending the oncology clinic at Tikur Anbessa specialized hospital.** *BMC Psychiatry* 2020, 20(1):446.
8. Khamseh ME, Baradaran HR, Javanbakht A, Mirghorbani M, Yadollahi Z, Malek M: **Comparison of the CES-D and PHQ-9 depression scales in people with type 2 diabetes in Tehran, Iran.** *BMC Psychiatry* 2011, 11(1):61.
9. APA: **Diagnostic and statistical manual of mental disorders (5th ed.);** 2013.
10. Negeri ZF, Levis B, Sun Y, He C, Krishnan A, Wu Y, Bhandari PM, Neupane D, Brehaut E, Benedetti A et al: **Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis.** *BMJ* 2021, 375:n2183.
11. Levis B, Benedetti A, Ioannidis JPA, Sun Y, Negeri Z, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D et al: **Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis.** *J Clin Epidemiol* 2020, 122:115-128.e111.
12. Manea L, Boehnke JR, Gilbody S, Moriarty AS, McMillan D: **Are there researcher allegiance effects in diagnostic validation studies of the PHQ-9? A systematic review and meta-analysis.** *BMJ Open* 2017, 7(9):e015247.
13. Levis B, Benedetti A, Levis AW, Ioannidis JPA, Shrier I, Cuijpers P, Gilbody S, Kloda LA, McMillan D, Patten SB et al: **Selective Cutoff Reporting in Studies of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool.** *Am J Epidemiol* 2017, 185(10):954-964.
14. Malpass A, Shaw A, Kessler D, Sharp D: **Concordance between PHQ-9 scores and patients' experiences of depression: a mixed methods study.** *Br J Gen Pract* 2010, 60(575):e231-238.
15. Malpass A, Dowrick C, Gilbody S, Robinson J, Wiles N, Duffy L, Lewis G: **Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study.** *Br J Gen Pract* 2016, 66(643):e78-84.
16. Katz IR, Liebmann EP, Resnick SG, Hoff RA: **Performance of the PHQ-9 across conditions and comorbidities: Findings from the Veterans Outcome Assessment survey.** *J Affect Disord* 2021, 294:864-867.

-
17. Kroenke K, Spitzer RL, Williams JB, Löwe B: **The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review.** *Gen Hosp Psychiatry* 2010, **32**(4):345-359.
 18. Zimmerman M, Martinez JH, Friedman M, Boerescu DA, Attiullah N, Toba C: **How can we use depression severity to guide treatment selection when measures of depression categorize patients differently?** *J Clin Psychiatry* 2012, **73**(10):1287-1291.
 19. Levis B, Benedetti A, Thombs BD: **Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis.** *Bmj* 2019, **365**:l1476.
 20. Hansson M, Chotai J, Nordstöm A, Bodlund O: **Comparison of two self-rating scales to detect depression: HADS and PHQ-9.** *Br J Gen Pract* 2009, **59**(566):e283-288.
 21. Joffres M, Jaramillo A, Dickinson J, Lewin G, Pottie K, Shaw E, Connor Gorber S, Tonelli M: **Recommendations on screening for depression in adults.** *Cmaj* 2013, **185**(9):775-782.
 22. Newson J, Hunter D, Thiagarajan T: **The Heterogeneity of Mental Health Assessment.** *Front Psychiatry* 2020, **11**:76.
 23. Newson JJ, Pastukh V, Thiagarajan TC: **Poor Separation of Clinical Symptom Profiles by DSM-5 Disorder Criteria.** *Frontiers in Psychiatry* 2021, **12**.
 24. Lamers F, van Oppen P, Comijs HC, Smit JH, Spinoven P, van Balkom AJ, Nolen WA, Zitman FG, Beekman AT, Penninx BW: **Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA).** *J Clin Psychiatry* 2011, **72**(3):341-348.
 25. Flory JD, Yehuda R: **Comorbidity between post-traumatic stress disorder and major depressive disorder: alternative explanations and treatment considerations.** *Dialogues Clin Neurosci* 2015, **17**(2):141-150.
 26. Newson J, Thiagarajan T: **Assessment of Population Well-Being With the Mental Health Quotient (MHQ): Development and Usability Study.** *JMIR Ment Health* 2020, **7**(7):e17935.
 27. Newson J, Pastukh V, Thiagarajan T: **Assessment of Population Well-being With the Mental Health Quotient: Validation Study.** *JMIR Mental Health* 2022, **9**(4):e34105.
 28. Newson J, Thiagarajan T: **Dynamic Dataset of Global Population Mental Wellbeing.** *PsyArXiv* 2021.
 29. <https://sapienlabs.org/mhq/>
 30. Schneider D, Harknett K: **What's to Like? Facebook as a Tool for Survey Data Collection.** *Sociological Methods & Research* 2022, **51**(1):108-140.
 31. 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.
 32. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P: **Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders.** *The American Journal of Psychiatry* 2010, **167**(7):748-751.
 33. Cuthbert BN, Insel TR: **Toward the future of psychiatric diagnosis: the seven pillars of RDoC.** *BMC Medicine* 2013, **11**(1):126.
-

-
34. Bianchi R, Verkuilen J, Toker S, Schonfeld IS, Gerber M, Brähler E, Kroenke K: **Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study.** *Psychol Assess* 2022, **34**(6):595-603.
35. Fried EI: The 52 symptoms of major depression: **Lack of content overlap among seven common depression scales.** *Journal of Affective Disorders* 2017, **208**:191-197.
36. Beeker T, Mills C, Bhugra D, Te Meerman S, Thoma S, Heinze M, von Peter S: **Psychiatrization of Society: A Conceptual Framework and Call for Transdisciplinary Research.** *Front Psychiatry* 2021, **12**:645556.
37. Wakefield JC: Diagnostic Issues and Controversies in DSM-5: **Return of the False Positives Problem.** *Annu Rev Clin Psychol* 2016, **12**:105-132.
38. Lalji HM, McGrogan A, Bailey SJ: **An analysis of antidepressant prescribing trends in England 2015-2019.** *J Affect Disord Rep* 2021, **6**:100205.
39. Groen RN, Ryan O, Wigman JTW, Riese H, Penninx B, Giltay EJ, Wichers M, Hartman CA: **Comorbidity between depression and anxiety: assessing the role of bridge mental states in dynamic psychological networks.** *BMC Med* 2020, **18**(1):308.
40. Hirschfeld RM: The Comorbidity of Major Depression and Anxiety Disorders: **Recognition and Management in Primary Care.** *Prim Care Companion J Clin Psychiatry* 2001, **3**(6):244-254.
41. Goldberg D: **The heterogeneity of «major depression».** *World Psychiatry* 2011, **10**(3):226-228.
42. 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.
43. Zimmerman M, Ellison W, Young D, Chelminski I, Dalrymple K: **How many different ways do patients meet the diagnostic criteria for major depressive disorder?** *Compr Psychiatry* 2015, **56**:29-34.
44. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, Fan CC, Degenhardt L, Ganna A, Greve AN et al: **Exploring Comorbidity Within Mental Disorders Among a Danish National Population.** *JAMA Psychiatry* 2019, **76**(3):259-270.
45. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, Harrington H, Hogan S, Poulton R, Ramrakha S et al: **Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study.** *JAMA Netw Open* 2020, **3**(4):e203221.
46. Nandi A, Beard JR, Galea S: **Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review.** *BMC Psychiatry* 2009, **9**:31.
47. Harris MG, Kazdin AE, Chiu WT, Sampson NA, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Altwaijri Y, Andrade LH, Cardoso G et al: **Findings From World Mental Health Surveys of the Perceived Helpfulness of Treatment for Patients With Major Depressive Disorder.** *JAMA Psychiatry* 2020, **77**(8):830-841.
48. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A et al: **Erratum: Resting-state connectivity biomarkers define neurophysiological subtypes of depression.** *Nat Med* 2017, **23**(2):264.
-

PHQ-9, Life Impact, and Breadth of Symptomatic Experience

SUPPLEMENTARY MATERIALS

Supplementary Table 1:

DEMOGRAPHIC BREAKDOWN OF THE SAMPLE BY AGE, GENDER, AND COUNTRY.

	N	%
AGE GROUP		
18-24	417	7.3%
25-34	554	9.6%
35-44	722	12.6%
45-54	1030	17.9%
55-64	1405	24.4%
65-74	1137	19.8%
75-84	436	7.6%
85+	49	0.9%
BIOLOGICAL SEX		
Male	2451	42.6%
Female	3257	56.6%
Other	42	0.7%
COUNTRY		
United States	967	16.8%
United Kingdom	593	10.3%
Nigeria	555	9.6%
India	548	9.5%
Pakistan	443	7.7%
Philippines	413	7.2%
South Africa	331	5.7%
Trinidad and Tobago	262	4.5%
Canada	251	4.3%
Australia	230	4.0%
Sri Lanka	169	2.9%
Ireland	157	2.7%
Zimbabwe	141	2.4%
New Zealand	138	2.4%
Malaysia	119	2.1%
Ghana	112	1.9%
Other countries	344	6.0%

Supplementary Table 2:

T-TEST (P VALUES) SHOWING STATISTICAL SIGNIFICANCE BETWEEN LIFE IMPACT RATINGS FOR DIFFERENT MHQ ITEMS (THAT WERE 1-1 MATCHED TO PHQ-9 ITEMS). FOR PHQ-9 RATINGS OF "NEARLY EVERY DAY".

Spectrum items:

	SLEEP QUALITY	ENERGY LEVEL	FOCUS & CONCENTRATION	APPETITE REGULATION
SLEEP QUALITY	NA	$p = 1.54E-34$	$p = 8.63E-26$	$p = 1.29E-09$
ENERGY LEVEL	NA	NA	$p = 0.06862$	$p = 3.43E-07$
FOCUS & CONCENTRATION	NA	NA	NA	$p = 2.43E-08$
APPETITE REGULATION	NA	NA	NA	NA

Problem Items:

FEELINGS OF SADNESS, DISTRESS OR HOPELESSNESS	
Suicidal thoughts or intentions	$p = 5.26E-02$

Supplementary Table 3:
 STATISTICAL SIGNIFICANCE (TUKEYS &
 ANOVA) OF DIFFERENCES IN THE AVERAGE
 MHQ LIFE IMPACT RATINGS OF EACH ITEM
 CORRESPONDING TO EACH SUCCESSIVE
 PHQ-9 RATING SELECTION.

PHQ/MHQ ITEM	COMPARISON BETWEEN SUCCESSIVE PHQ_9 RATING SELECTION	DIFFERENCE IN THE AVERAGE MHQ LIFE IMPACT RATING	p VALUE (TUKEYS)	p VALUE (ANOVA)
Trouble falling or staying asleep, or sleeping too much / Sleep quality	1-0	-1.765875215	0.00E+00	0
	2-1	-0.907113485	0	0
	3-2	-1.243533716	0	0
Feeling tired or having little energy / Energy level	1-0	-1.209980781	0	0
	2-1	-0.92108413	0.00E+00	0
	3-2	-1.282801028	0.00E+00	0
Trouble concentrating on things, such as reading the newspaper or watching television / Focus & concentration	1-0	-1.340065151	0	0
	2-1	-0.952768265	0	0
	3-2	-0.52086574	3.75E-05	0
Poor appetite or overeating / Appetite regulation	1-0	-1.456708499	0.00E+00	0
	2-1	-0.933871964	0	0
	3-2	-0.973177743	0	0
Thoughts that you would be better off dead, or of hurting yourself in some way / Suicidal thoughts or intentions	1-0	2.634436925	0	0
	2-1	1.573308837	0.00E+00	0
	3-2	1.286403803	0.00E+00	0
Feeling down, depressed, or hopeless / Feelings of sadness, distress or hopelessness	1-0	2.224389132	0	0
	2-1	1.384730282	0	0
	3-2	1.606134569	0	0

Supplementary Figure 1 - PROPORTION OF RESPONDENTS FOR EACH LIFE IMPACT RATING SCORE (1-9) ACROSS PHQ-9 SUM SCORE CATEGORIES FOR THE MATCHED ITEM OF:

