# Reliability and Validity of the Mental Health Quotient (MHQ)

**Jennifer J Newson, PhD[1]\* Vladyslav Pastukh[1], Tara Thiagarajan, PhD[1]**

[1]Sapien Labs, 1201 Wilson Blvd, 27th floor, Arlington, VA, United States of America

**\* Correspondence:**
Jennifer Jane Newson, PhD
Email: jennifer@sapienlabs.org

## Abstract

***Background:***

The MHQ is an assessment of mental health and wellbeing that comprehensively covers symptoms across 10 major psychiatric disorders as defined by the DSM-5, in addition to constructs defined by RDoC and positive dimensions of mental function using a novel life-impact scale. An overall measure of mental wellbeing, the *Mental Health Quotient* or MHQ, is computed based on these elements using a nonlinear transformation of the scale followed by a rescaling. The MHQ has been deployed as part of the Mental Health Million Project as a freely available anonymous online assessment that, on completion, provides a score to the individual that places them on a spectrum from Distressed to Thriving along with a personal report spanning their various dimensions of mental wellbeing with strategies for improvement. Since its launch in April 2020 over 200,000 people have taken the MHQ. Here we provide various demonstrations of the reliability and validity of the MHQ.

***Objective:***

This paper outlines the reliability and validity of the Mental Health Quotient (MHQ), including construct

validity of the life impact scale, sample and test-retest reliability of the assessment and criterion validation of the MHQ with respect to productivity loss and clinical burden.

### *Methods:*

To assess  sample reliability, random demographically matched samples of 11,033 people were compared from within the same 6-month period. Test-retest reliability was determined using the subset of individuals who had taken the assessment twice at least 3 days apart (N=1907). In addition, a subset of respondents (N=4,247 or 7,625) were asked additional questions (along with the standard MHQ questions) on symptom frequency and severity for an example symptom (Feelings of Sadness, Distress or Hopelessness), days of work missed in the past month, and days with reduced productivity. In addition, elements with high negative life impact considered to meet the threshold to be considered a 'symptom' were mapped to the criteria for each of 10 major DSM-5 based mental health disorders to calculate the clinical burden (N=174,618).

### *Results:*

Distinct samples collected during the same period had indistinguishable MHQ distributions and average ratings for each of the 47 elements, demonstrating the reliability of the assessment and MHQ scores were correlated with r=0.84 between retakes. The life impact rating was correlated with both frequency and severity of symptoms and mean values had a clear linear relationship with an $R^2>0.99$. Furthermore, aggregate MHQ scores were systematically related to both productivity and clinical burden. At one end of the scale, those in the Distressed category had an average productivity loss of 15.2±0.5 days per month with 89.08% (8,986/10,087) mapping to 1 or more DSM-5 based clinical disorders. In contrast those at the other end of the scale, in the Thriving category, had an average productivity loss of 1.3±0.1 and 0.00% (1/24,365) had any DSM-5 based clinical disorder.

### *Conclusions:*

The MHQ is a valid and reliable assessment of mental wellbeing when delivered anonymously online.

# 1. Introduction

## 1.1 Background

The landscape of mental health assessment is built around specific psychiatric disorder categories based on the clinical classification systems of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [1] or the International Classification of Diseases (ICD-11) [2]. Although these assessments can identify whether an individual exhibits symptoms pertaining to a specific mental health disorder definition, they do not provide a perspective of the overall symptom profile of the individual or insight into the positive assets and strengths of the individual's mental function. Thus, no assessment existed that could provide a perspective on the overall mental wellbeing of an individual across a general population where individuals fall along a continuum from distressed to thriving.

## 1.2 The Mental Health Quotient

To address this gap, we developed a new web-based assessment tool called the Mental Health Quotient (MHQ) [3], that assesses the complete breadth of mental health elements spanning a range from negative symptoms to positive mental assets using a unique life-impact scale. The tool was developed based on a comprehensive review of symptoms assessed across 126 commonly used psychiatric assessment tools spanning the disorders of depression, anxiety, bipolar disorder, attention-deficit-hyperactivity disorder (ADHD), post-traumatic stress disorder (PTSD), obsessive-compulsive disorder (OCD), addiction, schizophrenia, eating disorder and autism spectrum disorder (ASD), as well as cross disorder tools (see [4] for a complete list of assessment tools). Symptoms from these

126 assessments were consolidated into a set of 43 symptom categories and reviewed and expanded in the context of the Research Domain Criteria (RDoC) constructs put forward by the National Institute of Mental Health (NIMH) [5, 6]. The resultant 47 items were then split into 2 formats – mental functions that could manifest as a spectrum from positive to negative (spectrum questions), and those symptoms that purely represented detractions from overall mental health (problem questions).

Spectrum and problem questions within the MHQ were answered using a 9-point scale reflecting the consequence on one's life functioning and impact on their ability to carry out tasks and activities in their daily lives. The scale is therefore different from traditional mental health assessments which typically focus on the frequency, severity, duration and/or timing of symptoms. An aggregate MHQ score, developed using an algorithm that nonlinearly transforms the life impact scale based on different categories of symptom seriousness was provided on completion of the assessment [3]. This score is intended as a representation of the overall mental wellbeing of the individual and categorized from Distressed (-100) to Thriving (+200).

## 1.3 Clinical Burden and Productivity

The MHQ is currently used online as part of an open data project called the Mental Health Million Project, which is a web-based surveillance platform monitoring the status of population mental health across the globe, and currently spans 30 countries and 3 languages (English, French and Spanish). One purpose of the project is to understand the overall clinical burden of mental distress across the globe. For example, evidence suggests that the clinical burden of mental health has been increasing as many

individuals grapple with the consequences of the Covid-19 pandemic [7-9]. Therefore, mapping the scale of this mental health burden with a comprehensive symptom profiling measure that is valid, reliable and applicable to the general population is important to provide real-time, dynamic surveillance of population mental health.

A second purpose of the project is to better understand the consequences of poor mental wellbeing on life functioning outcomes such as productivity. Evidence suggests a clear relationship between mental wellbeing and productivity, where those with mental health disorders such as depression experience more days lost and unproductive days (e.g., presenteeism) compared to those without mental health disorders [10-16]. The significant scale of economic loss associated with this reduced productivity [17-20] means it is imperative that we better understand the relationship between mental wellbeing and productivity in the general population to develop preventative interventions, as well as having the necessary insights to help develop therapies and treatments that improve productivity in those already experiencing mental distress, therefore mitigating economic loss [21].

Altogether, it is important that an assessment of mental wellbeing across a population serves as a reliable and valid indicator of its mental health and productivity. Here we demonstrate the validity and reliability of the MHQ by addressing 1) how the unique life impact scale relates to more commonly used assessments of frequency and severity, 2) whether an anonymous web-based assessment serves as a true measure of the population by demonstrating its population and test-retest reliability, 3) how well the composite MHQ score relates to functional criteria such as clinical diagnostic criteria, workdays missed and overall life productivity.

# 2. Methods

## 2.1 Recruitment of Participants

The data used was taken from the Mental Health Million open-access database [22] and included responses from 179,238 English speaking individuals from the United States, Canada, United Kingdom, Ireland, Australia, New Zealand, South Africa, Singapore, India and Nigeria collected during 2021. The study received ethics approval from Health Media Lab Institutional Review Board. Participants were recruited via outreach campaigns on Facebook and using Google Ads by targeting a broad cross section of adults aged 18-85+ across a wide geographic and socioeconomic demographic. The anonymous assessment was freely available online for anyone to complete and individuals took the assessment for the purpose of getting their personalized mental wellbeing report on completion. These factors therefore motivated respondents to complete the assessment with honest and accurate answers. Only respondents who found the assessment easy to understand (i.e., responded *Yes* to the question, "Did you find this assessment easy to understand?") were included in the analysis. This resulted in the exclusion of 2.58% (174,618/179,238) respondents leaving 174,618 for the full analysis.

## 2.2 Assessment of Reliability

### 2.2.1 Reliability across randomly selected samples

All respondents from the United States, India, Australia and United Kingdom between January 2021and June 2021 were pooled together (N= 44,132). 4 randomly selected and non-overlapping samples of 11,033 people were selected with similar demographic composition. We then computed the average rating (1-9) for each MHQ scored element for each sample,

the average MHQ score for each sample and the statistical differences between the samples.

### 2.2.2 Internal consistency analysis

The MHQ is designed to be as parsimonious as possible without repetition. However, the internal consistency of the MHQ was evaluated by looking at the relative correlations between elements that would be expected to be correlated compared to those that would not. Firstly, the correlation between two questions asking about sleep quality within the MHQ was computed. Sleep question 1 asked respondents to "Assess your Sleep Quality: Your ability to easily fall asleep, stay asleep during the night and wake up feeling rested" and sleep question 2 asked respondents: "In general, I get as much sleep as I need". The 1-9 rating score from sleep question 1 was correlated against transformed answers from sleep question 2 where each answer option was assigned a number that was roughly equivalent to the text description: All the time = 7; Most of the time = 5; Some of the time = 3; Hardly ever = 1. Secondly, the correlation between two questions asking about mood was computed. Mood question 1 asked respondents to: "Assess your Feelings of Sadness, Distress or Hopelessness: Experiencing overwhelming feelings of sadness, distress or helplessness, or having spells of uncontrollable crying" and was rated on a 1-9 life impact scale. Mood question 2 asked respondents: "How would you describe your overall mood right now on a scale from very negative to very positive?" and was rated on a 1-9 scale from "Very negative" to "Very positive". Comparisons were also made between responses for the related MHQ items of Self Worth and confidence to Self-image. In addition, comparisons were made between items of Physical intimacy to Memory; Emotional control to Coordination; and Memory to Emotional control which would not be expected to have significant correlation.

### 2.2.3 Test-retest reliability

Within the sample of 174,618 respondents, email addresses were provided by 80,955 respondents to receive their MHQ report. These email addresses were automatically converted into anonymous unique identifiers which allowed us to identify repeat respondents. Of these, 2231 respondents had taken the MHQ twice at varying time intervals up to 15 months from the time of the first assessment. Those who took the MHQ twice within the same day or immediately on the next day were excluded as they were more likely to be experimenting with answer choices than evaluating their own change over time in an honest way. Thus only those who had at least three days between attempts were included in the analysis (N=1907). We examined the test-retest reliability of the MHQ in this sample by looking at the correlation between the item ratings on the first and second attempt as well as the correlation between MHQ scores across both attempts.

### 2.3 Validation of the Life Impact Scale

The scored questions of the MHQ are formulated to be answered on a 9-point life-impact scale reflecting the impact of a particular mental aspect on one's ability to function [3]. For example, for questions pertaining to mental health challenges, 1 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". For the purpose of validation, for the question which asked individuals to rate the impact of their Feelings of Sadness, Distress or Hopelessness on this 9 point scale, 2 additional questions were asked when a value of 5 or more was selected (N=4,247): (1) "How many days in the last week did you experience these feelings" with options for selection from 0-7 (similar

to the format in depression screening tools such as the Center for Epidemiologic Studies Depression (CES-D) scale [23, 24] and (2) "On these days, how did these feelings impact your ability to function in life" with 5 options of increasing severity as follows coded from 1 (least severe) to 5 (most severe): 1= "They would come and go while I went about my life as normal"; 2= "I did what I had to do, but they were always there in the back of my mind"; 3= "I managed but it took extreme effort"; 4 = "They stopped me doing the things I usually do, or would want to do"; 5 = "They consumed me so much I was unable to get out of bed". The average frequency and severity and standard error (SEM) were then computed for each selection on the life impact scale from 5 to 9.

## 2.4 Relationship of MHQ Score to Clinical Burden

The computation of the MHQ score takes into account the number of severe symptoms (i.e. high negative life impact). Thus, the number of elements with a rating that signifies a high negative life impact decreases as MHQ score increases although the nonlinear weighting differs for different types of symptoms. Nonetheless, to assess how effectively the MHQ score relates to clinical burden, we mapped symptoms of the MHQ to the criterion for each of 10 major DSM-5 disorders (depression, anxiety, bipolar disorder, ADHD, PTSD, OCD, addiction, schizophrenia, eating disorder and ASD) and examined (1) the percentage of individuals meeting the diagnostic criteria for at least 1 disorder and (2) the average number of diagnoses per person for each MHQ score bin where the bin size was 25 MHQ points. The mapping of clinical diagnosis was done as follows:

For each of the 47 MHQ items, responses were determined to be clinically significant symptoms if they met a particular threshold of impact on the

individual's ability to function (hereafter referred to as "clinical symptom(s) or symptoms"). For problem items which represented a unidimensional scale of symptom severity from 1 to 9, the threshold selection for a clinical symptom was ≥8. For spectrum items where elements of mental function could be either a negative symptom or a positive asset (e.g. memory) and the 1 to 9 scale ranged from negative symptom (1-4) to positive asset (6-9), the threshold selection for a clinical symptom was a rating of ≤1.

MHQ items were first mapped to the symptoms within each of 10 common DSM-5 disorder criteria based on the closest semantic match (Newson et al, under review). As MHQ questions were formulated from DSM-based questions, all symptoms from the DSM-5 disorder criteria had an MHQ match. The specific criteria rules of the DSM-5 were then applied to arrive at a disorder diagnostic mapping (Newson et al, under review). For example, for depression a positive diagnosis requires that the individual has experienced at least 5 of the criteria with 1 of them being a depressed mood or loss of interest or pleasure. A set of rules using combinations of the DSM-5-mapped MHQ items was developed to align with these criteria descriptions for each of the 10 disorders. For each respondent (N=174,618), these rules were applied to their MHQ clinical symptom profile to determine diagnostic match to each of the 10 disorders. Although this does not categorically state that the individual would be diagnosed with that disorder in the context of a clinical interview, it indicates that their pattern of clinical symptoms broadly aligned with the diagnostic criteria for that disorder. However, we note the caveat that in bipolar disorder, symptoms denoted extreme versions of positive assets (e.g., *grandiosity and decreased need for sleep*) which were not fully articulated within the MHQ, while for OCD the MHQ items were broader

(e.g., obsessive thoughts were incorporated within a general item reflecting strange, unwanted and obsessive thoughts). Further, specific criterion of symptom timing was not included as this is not included in the MHQ which assesses the individual's current perception.

### 2.5 Relationships of MHQ Score to Productivity Criterion.

To assess the relationship between the MHQ score and measures of functional productivity a subset of participants (N=7,625) were asked 2 additional questions: (1) "How many days during the past month were you totally unable to work or carry out your normal activities because of problems with your physical or mental health" and (2) "How many days during the past month were you able to work and carry out your normal activities, but could not get as much done because of problems with your physical or mental health?" with options to select a number between 0 and 31. Individuals were then grouped by MHQ score in bins of 25 MHQ points and the average and SEM of days or work missed (M) and days with reduced productivity (R) were then computed for each bin. We then computed overall loss of life productivity for each individual as M+n*R where n represented the loss of productivity on those days and was assumed to range from 20% to a 50% loss of productivity. Data was examined for all respondents together, and for a subset of respondents (N=3,306) who responded "Employed/Self-Employed" to the MHQ question "Please select

which best describes your occupational status?" (alternative answer options included Homemaker, Unemployed, Retired, Studying or Not able to work).
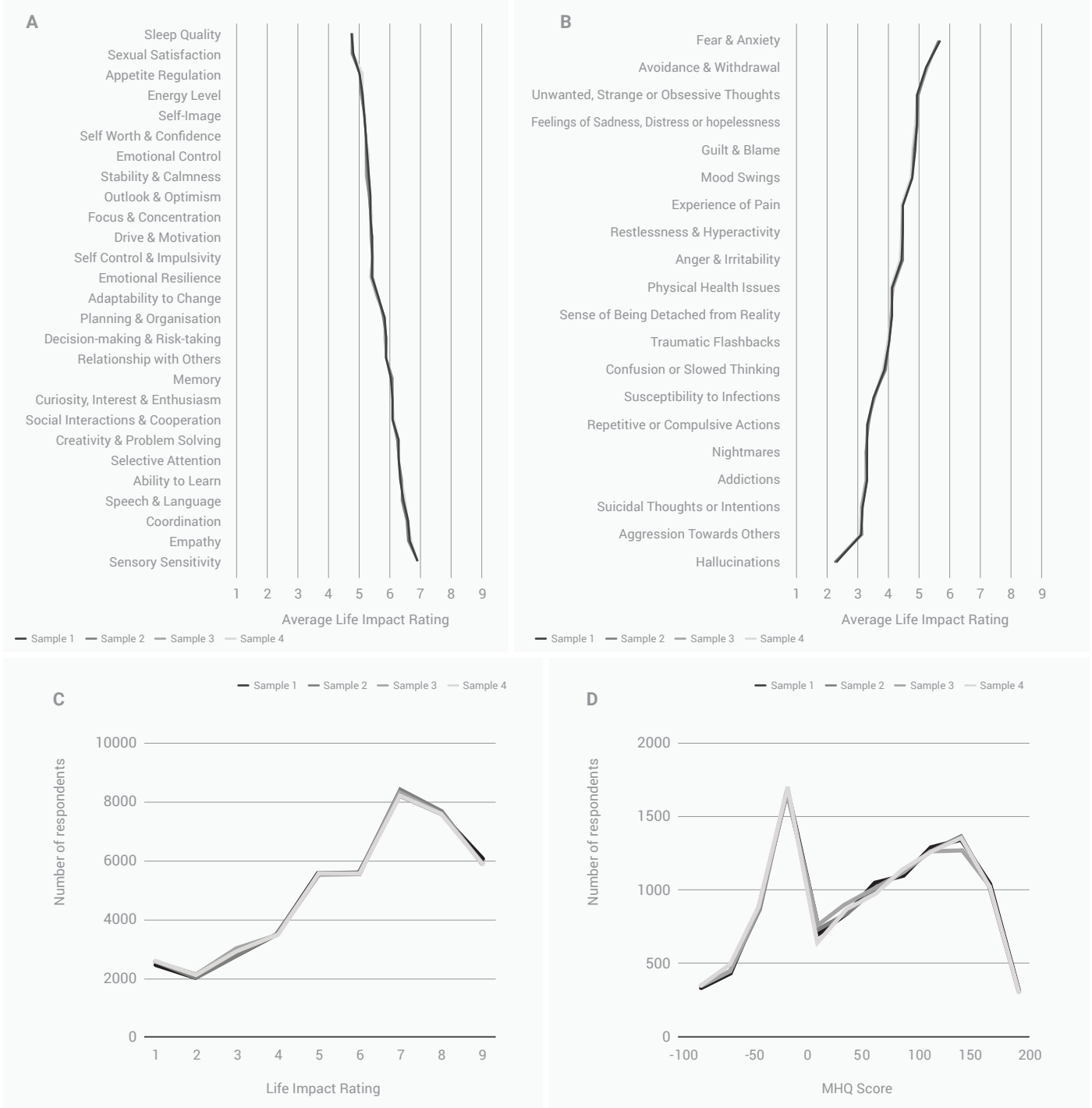
## 3. Results

### 3.1 Reliability and Internal Consistency in the MHQ

#### 3.1.1 Assessment Reliability

For a test to be reliable it must produce similar results across similar samples. Here we assessed the results across multiple randomly selected demographically matched samples of 11,033 people out of 44,132 respondents between January and June 2021 (see Methods). Figure 1A shows the average rating for each element of the assessment for spectrum (left) and problem (right) items. Across all samples the ratings were correlated with r>0.8 for all pairs and distributions of ratings for individual elements were highly similar and statistically indistinguishable (ANOVA, *P*=.99) with one example from the element "Self-image" shown in Figure 1C. Similarly, the distributions of the resulting MHQ scores for each of these 4 samples were highly similar (Figure 1C, ANOVA, *P*=.18).

These results confirm the reliability of the MHQ, offered anonymously and online. Should responses be randomly generated (e.g. by bots) or if individuals had highly inconsistent interpretations of the life impact scale this would not be the case.

**Figure 1: Reliability of the MHQ**

*(A) Average ratings of each of the 27 MHQ spectrum elements (left) and 20 MHQ problem items (right) in 4 separate samples of the MHQ obtained over a similar time period are indistinguishable (each bar is a sample).*

*(B) Distribution of ratings for one example MHQ element in each of the 4 samples (each line is a sample).*

*(C) Distribution of MHQ scores in each of the 4 samples (each line is a sample).*

### 3.1.2 Internal Consistency

Overall, related elements were more significantly correlated than unrelated elements examined. Among related elements, the life impact rating relating to *Sleep Quality* had a 0.63 correlation with response to the question on how many days a week they got sufficient sleep. Thus, those who had challenges with sleep quality were also likely to get fewer days of sufficient sleep. Similarly the life impact rating of *Feelings of sadness, distress or hopelessness* had a 0.64 correlation with momentary assessment of mood indicating that those with more significant impact of Feelings of sadness, distress or hopelessness were also more likely to have a negative mood at the time of taking the assessment. Finally, life impact rating of MHQ item *Self-image* had a 0.77 correlation with the rating of item *Self-worth and confidence.*
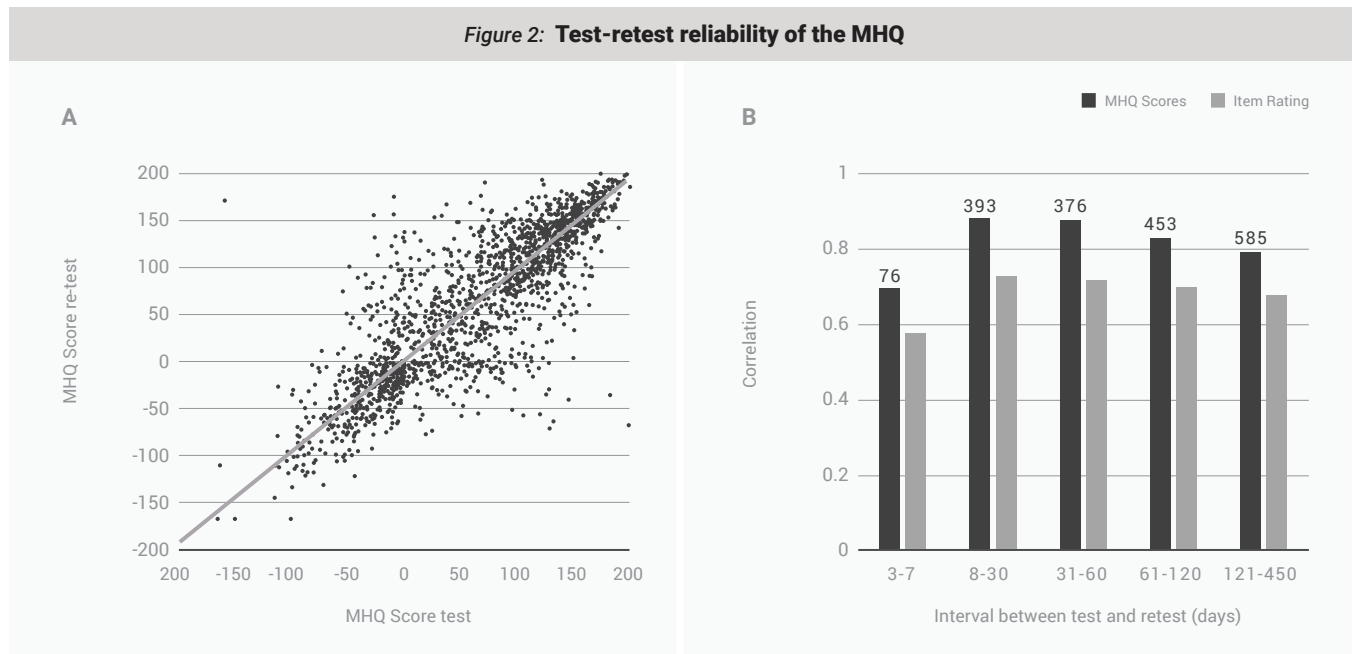
In contrast ratings of unrelated elements had lower

correlations. For example, *Memory* and *Physical intimacy* had a correlation of 0.35, *Emotional control* and *Coordination* had a correlation of 0.36 and *Memory* and *Emotional control* a correlation of 0.39.

### 3.1.3 Test-retest reliability

The MHQ is designed to measure changes in the mental wellbeing of the population and therefore in individual mental wellbeing status over time. Therefore MHQ scores of individuals could change over time. However over short time frames of less than a year, most individuals would not be expected to change significantly. In alignment with this, among all those who could be identified as having taken the MHQ twice at least 3 days apart, MHQ scores were correlated with r=0.84 (p~0). Figure 2A shows that the MHQ scores for test plotted against the MHQ scores for retest, demonstrating that points fall around the line y=x. Furthermore, this correlation did not change significantly as the interval between

---

**Figure 2:** **Test-retest reliability of the MHQ**



**(A)** *MHQ score for the assessment retake (retest) versus MHQ score for the first take (test). Grey line represents y=x.*

**(B)** *Correlation of MHQ scores (black) and item ratings (grey) for different test-retest intervals. Values of the bars represent N values for that group.*
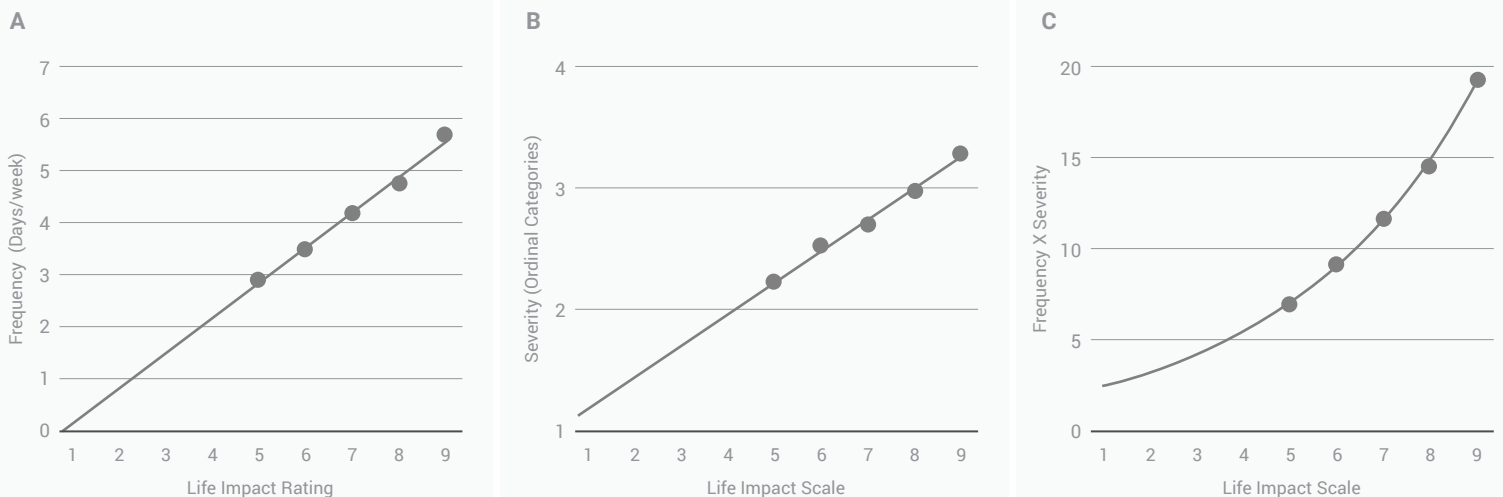
attempts increased although correlations were as high as 0.88 for retest intervals of 8-120 days (Figure 2B, black bars). Finally, the correlation between the ratings of individual elements on each attempt was 0.70 ($p<10^{-20}$) and did not change as the interval between attempts increased (Figure 2B, grey bars). Thus the MHQ had high test-retest reliability but also reflected changes that can occur in mental wellbeing over time.

### 3.2 Relationship of Life Impact Scoring to Frequency and Severity

Clinical assessments are heterogeneous in their assessment of the frequency and severity of symptoms. For example, a review of 126 assessment tools found that, across 19 commonly used depression scales, 51% of questions asked about frequency of symptoms and 32% asked about severity, while across 9 PTSD assessment tools,

17% of questions asked about frequency while 53% asked about severity [4]. Given the lack of clear understanding of which aspect – frequency, severity or duration of a symptom matters most, the MHQ uses a composite 9-point life impact scale as described in the methods. Here we demonstrate a strong relationship between the rating on the life impact scale and both frequency and severity of symptoms. In this example, individuals who indicated any degree of challenge (selection >4 on the 9-point life impact scale) with Feelings of sadness, distress and hopelessness were asked about the frequency of episodes as well as the nature of severity of the symptom from a "Feeling that comes and goes while they are going about their day" to "Being so consumed by it that they are unable to get out of bed". The rating of life impact was correlated to frequency, measured as the number of days in the past week where they experienced the symptom, with r=0.5. At an aggregate level mean ±SEM of frequency for



Figure 3: **Example relationship of Life Impact to Frequency and Severity**

**(A)** Selection on the MHQ life impact scale for Feelings of Sadness, Distress and Hopelessness is linearly related to frequency of these feelings or number of days per week where they were experienced (line = linear fit, $R^2$=0.99)

**(B)** Selection on the MHQ life impact scale is similarly linearly related to severity of the symptom (5 point ordinal scale; line = linear fit, $R^2$=0.99).

**(C)** Frequency x Severity is nonlinearly related to MHQ life impact selection (line = exponential fit, $R^2$ 0.98).
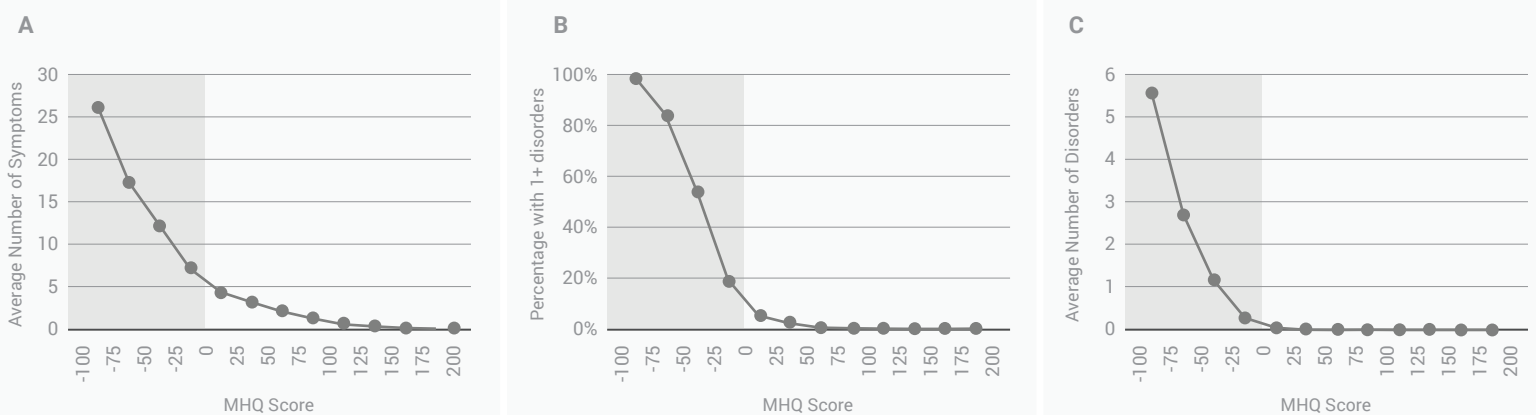
each rating on the life impact scale (data collected only for ratings >4 on the 9-point scale) was linearly related with $R^2=0.99$ (Figure 3A). Extrapolation of this function to life impact ratings less than 5 shows that those selecting 1 (the lowest end of the scale indicating no impact) would have experienced that symptom at a frequency of <1 day in the last week. Across all data, life impact was similarly positively correlated to the level of severity selected (where options of severity were coded from 1-5) but less so than with frequency (r=0.32). However, aggregate mean ±SEM of severity for each life impact rating was also linearly related (Figure 3B, $R^2=0.99$). Finally, in the aggregate, a composite measure of frequency x severity was nonlinearly related to the life impact rating with $R^2=0.98$ (Figure 3C).

### 3.3  Relationship of MHQ Scores to Clinical Burden

We next demonstrate a strong relationship between MHQ scores and clinical burden. Here, since symptoms asked about in the MHQ are derived from DSM-based assessments, each MHQ item maps to the various symptoms of each of 10 major mental health disorders including depression, bipolar disorder, anxiety, OCD, PTSD, ADHD, addiction, schizophrenia, eating disorder and ASD. By applying the diagnostic criterion of the DSM-5, a diagnostic indication can be determined for each disorder for each individual (see Methods). As would be expected, the average number of clinical symptoms increased as the MHQ decreased given the non-linear weighting of symptom severity within the MHQ scoring (Figure 4A). Beyond this, we show the percentage of people with symptom profiles that aligned with DSM-5 defined disorder criteria increased as the MHQ score decreased, such that 89.08% (8,986/10,087) of those in the Distressed category (MHQ <-50) had symptom profiles that aligned with at least one clinical mental health disorder while 0.00% (21/70367) in the categories of Succeeding and Thriving (MHQ >100) had profiles that aligned with these DSM-5 defined disorder criteria (Figure 4B). Similarly, the number of

---

**Figure 4:** **Relationship of MHQ Score to Clinical Symptoms and Diagnosis**



**(A)** Average number of symptoms (life impact >7 for problem items and <2 for spectrum items) decreased as MHQ score increased. Grey area represents the negative side of the scale or MHQ score categories of Distressed and Struggling; all panels.

**(B)** 97.83% (3,926/4,013) of those with the lowest negative MHQ scores

(−100 to −75) mapped to at least 1 of 10 major clinical disorders and decreased systematically such that <1% in the positive score range mapped to a clinical disorder.

**(C)** Average number of disorders per person decreased with MHQ score. In the positive range average disorders per person was close to 0.
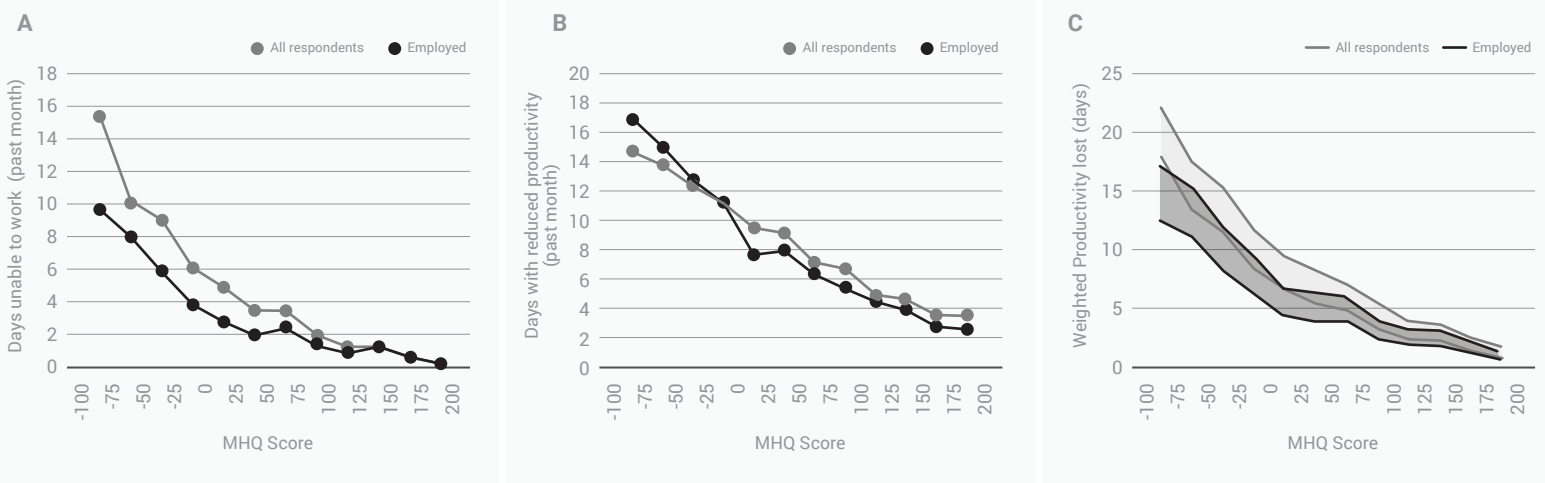
disorders per individual decreased systematically as MHQ scores increased with the average number of disorders per person at 3.8 for those in the Distressed group and 0.0 for those in the Succeeding and Thriving groups (Figure 4C). Thus, the MHQ score is also reflective of the overall clinical burden of mental health.

### 3.4 Loss of Function Criterion Validation

It is important that a score of mental wellbeing relevant to the general population reflects more tangible aspects of human function. We therefore explored the relationship between the MHQ score, as an overall assessment of mental wellbeing, to lost days of work and lost productivity (presenteeism). As MHQ scores increased, the average number of days of work missed in the past month (Figure 5A, mean ±SEM) decreased systematically and was best

fit by an exponential function with $R^2$=0.98. All those in the lowest MHQ score bin ($-75$ to $-100$) were unable to work 15.0±0.9 days while those who were employed (as opposed to studying, unable to work, unemployed, retired or occupied with household work) were unable to work 9.3±1.6 days in the last month. In contrast, those who were in the highest MHQ bin (175-200) lost only 0.2±0.1 days while those who were employed lost 0.2±0.07 days. Furthermore, as MHQ scores increased, average days where people reported not being as productive as usual at work (presenteeism) decreased linearly (Figure 5B, $R^2$=0.98 for all respondents and employed alone). Here, all those in the lowest MHQ bin ($-75$ to $-100$) were not productive 14.2±0.9 days while those employed alone were not productive 15.6 ±1.6 days which reduced to 3.2±0.8 and 2.2±1.1 days for those in the highest MHQ score bin (175-200) for all respondents and employed respondents respectively.

**Figure 5:** **Relationship of MHQ Score to Productivity**



*(A)* Days unable to work in the past month decreased nonlinearly as MHQ score increased (closed circles, exponential fit $R^2$ = 0.98). Employed people (open circles) with low MHQ scores missed fewer days of work on average compared to all respondents together.

*(B)* Days in the past month with reduced productivity (presenteeism) decreased linearly as MHQ score increased (closed circles, exponential fit $R^2$ = 0.98). Employed people (open circles) with low MHQ scores

had more days of presenteeism on average since they tended to miss fewer days of work.

*(C)* Total productivity loss for employed (dotted line) and all respondents together (solid line) as a function of MHQ score (calculated as Days missed + n * Days with reduced productivity, where n is assumed as a range between 0.2 [lower dotted/solid line] and 0.5 [upper dotted/ solid line]).

Figure 5C shows the total loss of life productivity as a function of MHQ score taking both days of work missed and days that were less productive, assuming a range of 20% to 50% loss of productivity on less productive days. Altogether those with the poorest mental wellbeing had an overall reduction in life productivity of anywhere from 18-23 days per month on average and while those with the best mental health did not often miss a day of work, even this group reported a few unproductive days a month. Thus, MHQ scores are a good representation of behavioral loss of function.

# 4. Discussion

Here we have demonstrated that the MHQ taken online has excellent sample and test-retest reliability and internal consistency, that the life impact scale used in the MHQ reflects a combination of both severity and frequency of symptoms and that the MHQ score relates systematically to clinical burden in the population as well as loss of function from the perspective of days of work missed and loss of productivity. For example, the results showed that the MHQ scores were highly similar and statistically indistinguishable between multiple randomly selected demographically matched samples of 11,033 people and MHQ scores were correlated between retakes with an r=0.84 . In addition, we showed that the correlation between MHQ scores and productivity outcomes of days lost and unproductive days (presenteeism) decreased systematically and was best fit by an exponential function with $R^2$=0.98, while the correlation between the frequency of an example symptom of Feelings of Sadness, Distress & Hopelessness and the MHQ life impact rating scale was r=0.5 with a linear relationships of $R^2$=0.99. Furthermore, the prevalence of MHQ-mapped disorders per individual decreased systematically

as MHQ score increased. Thus, the MHQ provides a valid and reliable estimation of population mental wellbeing that in turn reflects the clinical burden of mental health and productive capacity of a population. It is therefore also well suited to measure changes in the status and mental wellbeing of individuals and the overall population.

## 4.1 MHQ and Clinical Burden

Surveillance of population mental health requires assessments which are reliable, valid, accessible to the general population and provide a comprehensive profile of mental wellbeing which has clinical and real-world relevance. Currently, most assessments used in epidemiological studies are focused on the prevalence of individual disorders, rather than overall mental health and wellbeing. Thus, these tools treat mental illness as an all or nothing phenomenon rather than viewing mental wellbeing as a continuum from clinical distress to thriving [25]. In addition, focusing only on individual disorders creates a siloed landscape of clinical burden rather than offering a trans-diagnostic or disorder-agnostic perspective to mental distress that would overcome some of the difficulties of a system of disorder classification that is at odds with the heterogeneous and comorbid nature of symptomatic experience [26-33] and poorly articulated within the scope of existing psychiatric assessment tools [4, 34, 35]. Thus, the aggregate burden of clinical-level mental distress is generally unknown. The MHQ, as a valid and reliable measure of mental wellbeing, aims to fill this gap by providing an assessment tool that encompasses a full array of symptoms that can be mapped to multiple DSM-5 defined disorders and thus provides a useful estimate of both the prevalence of individual disorders and also the overall clinical distress or burden in the general population. When rolled out at scale, as is currently being actioned as part of the Mental Health Million

project, the MHQ provides a solid foundation for the global surveillance of population mental health.

### 4.2 MHQ and Productivity

Over the past few decades, there has been mounting evidence supporting the relationship between mental wellbeing and productivity [10-16], as well as the resultant economic loss to society as a consequence of days lost and unproductive days (e.g., presenteeism) [17-20]. With the increased prevalence in mental distress as a result of the Covid-19 pandemic [7-9], the need to better understand the relationship between mental wellbeing and productivity in the general population, and to develop interventions, treatments and workplace solutions that mitigate this loss, is ever more important. However, studies investigating this relationship rarely consider mental health and wellbeing as a continuum and use tools which are specific to certain disorders and that do not cover the breadth of symptomatic experience [4]. The systematic relationship observed between the spectrum of MHQ scores and productivity loss supports the use of the MHQ as an assessment of the productive capacity of a population, independent of any disorder classification. The MHQ therefore has important application in helping companies assess the mental wellbeing of their workforce and its impact on workplace productivity and healthcare costs. Similarly, it has important applications in assessing the mental wellbeing of University student bodies. This will allow, and encourage, organizations and institutions to be more strategic in their management of mental wellbeing.

# 5. Acknowledgements

# 6. Authors' Contributions

*JN and TT conceptualized, designed and led the study. JN and VP performed the analysis. JN and TT interpreted the data and drafted the manuscript. JN created the figures. TT and JN made critical revisions. JN, TT and VP approved the final version, and agreed to be accountable for all aspects of the work.*

# 7. Conflicts of Interest

*TT received a grant award from the National Institute of Mental Health (NIMH) to develop a commercial version of the MHQ tool referenced herein. There are no other commercial or financial relationships that could be construed as a potential conflict of interest.*

# 8. Abbreviations

**ADHD:** attention-deficit/hyperactivity disorder

**ASD:** autism spectrum disorder

**DSM:** *Diagnostic and Statistical Manual* of Mental Disorders

**ICD:** International Classification of Diseases

**MHQ:** Mental Health Quotient

**OCD:** obsessive-compulsive disorder

**PTSD:** post-traumatic stress disorder

**RDoC:** Research Domain Criteria

# 9.  References

**1.**  APA. Diagnostic and statistical manual of mental disorders (5th ed.)2013.

**2.**  WHO. International statistical classification of diseases and related health problems (11th Revision)2018.

**3.**  Newson JJ, Thiagarajan TC. Assessment of Population Well-Being With the Mental Health Quotient (MHQ): Development and Usability Study. JMIR Ment Health. 2020 2020/7/20;7(7):e17935. doi: 10.2196/17935.

**4.**  Newson JJ, Hunter D, Thiagarajan TC. The Heterogeneity of Mental Health Assessment. Frontiers in Psychiatry. 2020 2020-February-27;11(76). doi: 10.3389/fpsyt.2020.00076.

**5.**  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. American Journal of Psychiatry. 2010 2010/07/01;167(7):748-51. doi: 10.1176/appi. ajp.2010.09091379.

**6.**  Insel TR. The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry. American Journal of Psychiatry. 2014 2014/04/01;171(4):395-7. doi: 10.1176/appi. ajp.2014.14020138.

**7.**  Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. The Lancet Psychiatry. 2020;7(10):883-92. doi: 10.1016/S2215-0366(20)30308-4.

**8.**  Serafini G, Parmigiani B, Amerio A, Aguglia A, Sher L, Amore M. The psychological impact of COVID-19 on the mental health in the general population. QJM. 2020;113(8):531-7. PMID: 32569360. doi: 10.1093/qjmed/hcaa201.

**9.**  Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord. 2020 Dec 1;277:55-64. PMID: 32799105. doi: 10.1016/j.jad.2020.08.001.

**10.**  Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med. 1997 Jul;27(4):861-73. PMID: 9234464. doi: 10.1017/s0033291797004807.

**11.**  Alonso J, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, et al. Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys. Mol Psychiatry. 2011 Dec;16(12):1234-46. PMID: 20938433. doi: 10.1038/mp.2010.101.

**12.**  Burton WN, Schultz AB, Chen CY, Edington DW. The association of worker productivity and mental health: a review of the literature. International Journal of Workplace Health Management. 2008;1(2):78-94. doi: 10.1108/17538350810893883.

**13.**  Hemp P. Presenteeism: at work--but out of it. Harv Bus Rev. 2004 Oct;82(10):49-58, 155. PMID: 15559575.

**14.**  Lim D, Sanderson K, Andrews G. Lost productivity among full-time workers with mental disorders. J Ment Health Policy Econ. 2000 Sep 1;3(3):139-46. PMID: 11967449. doi: 10.1002/mhp.93.

**15.**  Evans-Lacko S, Knapp M. Global patterns of workplace productivity for people with depression:

absenteeism and presenteeism costs across eight diverse countries. Soc Psychiatry Psychiatr Epidemiol. 2016 Nov;51(11):1525-37. PMID: 27667656. doi: 10.1007/s00127-016-1278-4.

**16.** Bubonya M, Cobb-Clark DA, Wooden M. Mental health and productivity at work: Does what you do matter? Labour Economics. 2017 2017/06/01/;46:150-65. doi: https://doi.org/10.1016/j.labeco.2017.05.001.

**17.** Marcotte DE, Wilcox-Gök V. Estimating the employment and earnings costs of mental illness: recent developments in the United States. Soc Sci Med. 2001 Jul;53(1):21-7. PMID: 11386306. doi: 10.1016/s0277-9536(00)00312-9.

**18.** Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. Jama. 2003 Jun 18;289(23):3135-44. PMID: 12813119. doi: 10.1001/jama.289.23.3135.

**19.** Trautmann S, Rehm J, Wittchen H-U. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep. 2016;17(9):1245-9. PMID: 27491723. doi: 10.15252/embr.201642951.

**20.** Whiteford H. The Productivity Commission inquiry into mental health. Australian & New Zealand Journal of Psychiatry. 2021:00048674211031159. doi: 10.1177/00048674211031159.

**21.** Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. The Lancet Psychiatry. 2016;3(5):415-24. doi: 10.1016/S2215-0366(16)30024-4.

**22.** Newson J, Thiagarajan T. Dynamic Dataset of Global Population Mental Wellbeing (preprint). PsyArXiv (2021). doi: https://doi.org/10.31234/osf.io/vtzne.

**23.** Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement. 1977 1977/06/01;1(3):385-401. doi: 10.1177/014662167700100306.

**24.** 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, Volume 3, 3rd ed. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2004. p. 363-77.

**25.** Rössler W. What is Normal? The Impact of Psychiatric Classification on Mental Health Practice and Research. Front Public Health. 2013;1:68. PMID: 24350236. doi: 10.3389/fpubh.2013.00068.

**26.** Angold A, Costello EJ, Erkanli A. Comorbidity. Journal of Child Psychology and Psychiatry. 1999 1999/01/01;40(1):57-87. doi: https://doi.org/10.1111/1469-7610.00424.

**27.** Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG. Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. Br J Psychiatry Suppl. 1996 Jun(30):17-30. PMID: 8864145.

**28.** Maj M. 'Psychiatric comorbidity': an artefact of current diagnostic systems? British Journal of Psychiatry. 2005;186(3):182-4. doi: 10.1192/bjp.186.3.182.

**29.** Feczko E, Miranda-Dominguez O, Marr M, Graham

AM, Nigg JT, Fair DA. The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. Trends Cogn Sci. 2019 Jul;23(7):584-601. PMID: 31153774. doi: 10.1016/j.tics.2019.03.009.

**30.** Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. Journal of affective disorders. 2015;172:96-102. PMID: 25451401. doi: 10.1016/j.jad.2014.10.010.

**31.** Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study. JAMA Network Open. 2020;3(4):e203221-e. doi: 10.1001/jamanetworkopen.2020.3221.

**32.** McElroy E, Patalay P. In search of disorders: internalizing symptom networks in a large clinical sample. J Child Psychol Psychiatry. 2019;60(8):897-906. PMID: 30900257. doi: 10.1111/jcpp.13044.

**33.** 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.

**34.** Allsopp K, Read J, Corcoran R, Kinderman P. Heterogeneity in psychiatric diagnostic classification. Psychiatry Research. 2019 2019/09/01/;279:15-22. doi: https://doi.org/10.1016/j.psychres.2019.07.005.

**35.** Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. J Affect Disord. 2017 Jan 15;208:191-7. PMID: 27792962. doi: 10.1016/j.jad.2016.10.019.