Psychiatry Research 280 (2019) 112512

Contents lists available at ScienceDirect

Psychiatry Research

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

How should the severity of depression be rated on self-report depression scales?

Mark Zimmerman[a,b,*], Sophie Kerr[a,b]

[a] Department of Psychiatry and Human Behavior, Brown Medical School, Rhode Island Hospital, Providence, United States
[b] Department of Psychiatry, Rhode Island Hospital, Providence, United States

ARTICLE INFO

Keywords:
Severity
Depression
Symptom severity
Symptom frequency
Symptom intensity
Symptom scale

ABSTRACT

Almost all depression measures have been developed without discussing how to best conceptualize and assess the severity of depression. The most valid rating format of depression severity scales is unsettled and has been little studied. In the present study from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we compared the validity of alternative approaches towards rating the severity of depressive symptoms. Data was collected using TurkPrime. One hundred eighty-five participants currently in treatment for a mental health problem completed a self-report measure of depression and rated the symptoms on two 4-point ordinal scales assessing symptom frequency and symptom intensity. The respondents also rated their global level of depression severity and completed a measure of psychosocial functioning and quality of life. The symptom ratings based on intensity and frequency were highly correlated with each other, and equally highly correlated with subjects' global rating of overall severity of depression, as well as ratings of psychosocial functioning and quality of life. A composite index of severity based on the sum of frequency and severity ratings was no more highly correlated with the external validators. The results of the present study suggest that ratings of depressive symptoms based on either symptom intensity or symptom frequency are equally valid.

1. Introduction

The severity of depression has been most frequently quantified on paper-and-pencil and clinician-administered rating scales. There are numerous instruments assessing depressive symptoms, with variability amongst them in the time frame covered (the two most common time frames being the past one or two weeks), rating guidelines (most scales use Likert-type ratings based on symptom frequency, persistence or intensity), and item content.

Descriptions of scale construction typically focus on the content of the measure. Some measures are linked to the symptom criteria that are used to diagnose depression (Kroenke et al., 2001; Olsen et al., 2003; Rush et al., 2003; Zimmerman et al., 2008), whereas others assess a broad range of features that patients indicate are most important in measuring outcome (Zimmerman et al., 2011) or assess a range of diagnostic and associated symptoms of depression (Rush et al., 1996). In developing the Multidimensional Depression Assessment Scale, Cheung and Power (2012) reviewed the content of fifteen depression scales and how the scale they had developed would address a content gap. There was no discussion, however, of rating formats and why a symptom frequency format was chosen for their measure rather than a rating format based on symptom intensity. Rarely do scale developers discuss the reason for choosing a particular rating format.

Little research has examined which rating parameters provide the most valid indicator of depression severity. Is the severity of depression best conceptualized as the number of symptoms (i.e., present or absent), frequency of symptoms (e.g., every day vs. half the days vs. few days), persistence of symptoms (e.g., always present vs. often present vs. sometimes present), or intensity of symptoms (e.g., severe vs. moderate vs. mild)? Williams et al. (2008), in standardizing the scoring of the Hamilton Depression Rating Scale (HAMD), created a grid scoring format to incorporate information regarding symptom frequency/persistence and intensity in the ratings. In developing the Patient-Reported Outcomes Measurement Information System (PROMIS) depression scale, Pilkonis et al. (2011) reviewed studies comparing alternative response options and concluded that frequency scaling outperformed intensity ratings, though the studies reviewed were not studies of depression ratings.

The most valid rating format of depression severity scales is unsettled and has been little studied. We are aware of only one study comparing alternative rating formats in assessing depression. Parker and colleagues examined whether it is important to consider both

* Corresponding author at: 146 West River Street, Providence, RI 02904, United States.
E-mail address: mzimmerman@lifespan.org (M. Zimmerman).

https://doi.org/10.1016/j.psychres.2019.112512
Received 26 May 2019; Received in revised form 8 August 2019; Accepted 8 August 2019
Available online 09 August 2019
0165-1781/

intensity and frequency constructs and found that symptom intensity was a better indicator of severity than symptom frequency (Parker et al., 1994).

In the present study from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we compared the validity of alternative approaches towards rating the severity of depressive symptoms. We hypothesized that ratings based on both symptom intensity and symptom frequency would be more highly correlated with ratings of psychosocial impairment and quality of life than ratings based on symptom intensity alone or symptom frequency alone.

2. Methods

Data was collected using TurkPrime, an integration of research platforms that allows for efficient collection of large samples of data. TurkPrime manages a panel of Mechanical Turk respondents and partners with multiple other online market research platforms to achieve a global reach of over 15 million respondents. The respondents are offered the opportunity to complete surveys and other brief tasks online from the location of their choosing for compensation. The amount and type of compensation (e.g. cash, gift cards) depends on which online platform the participants were drawn from. Researchers were provided with a financial quote for data collection and TurkPrime managed the compensation of participants. The total cost was less than $8.00 per participant.

The respondent base of TurkPrime is referred to as the Prime Panel and participants must be 18 years old. TurkPrime integrates with the widely used Amazon Mechanical Turk (MTurk) as well as other online platforms and provides services to improve the quality and functionality of data collection. Unlike MTurk, TurkPrime offers participants the opportunity to provide demographic and behavioral information on an ongoing basis for no compensation and then uses this information to later recruit certain types of samples (e.g., Davidai, 2018; Skrzynski et al., 2018). This approach is preferable to studies that explicitly state inclusion or exclusion criteria in which participants are more likely to respond dishonestly in order to be able to participate in the study (Chandler and Shapiro, 2016).

For the current study, TurkPrime provided researchers with a financial quote to recruit and compensate a sample of 180 participants currently in treatment for a mental health problem. TurkPrime assessed current psychiatric treatment as one of their ongoing pre-screening questions using the statement, “Are you currently in treatment for a mental health problem?” Those who endorsed current psychiatric treatment were presented with a link to complete our survey on Qualtrics Survey Software as one of the available tasks on their online dashboard. After following the link, participants were given the opportunity to consent. Those who refused consent were navigated away from the study. Participants then completed the survey which included demographic questions, questions regarding their psychiatric treatment history, a depression scale with both intensity and frequency ratings of symptoms, and scales measuring functioning and quality of life. The survey concluded with a debriefing message including the research coordinator's email to contact with questions and concerns as well as resources to locate mental health professionals in their area.

The survey included 3 ‘attention check’ questions as a measure of validity to ensure that participants were reading and responding to each item. The attention check questions followed the same format as the other survey question but directed the respondent to select a specific answer. For example, the attention check item embedded in the depression scale frequency items stated “Please select 0 days.” There was no symptom listed for this item, rather just the instruction to select 0 days. Analogously, on the intensity subscale the respondent was instructed to select “not present” for an item where a symptom was not listed. Following the depression scale an item was included which again simply directed the respondent to select a particular response option. Participants were excluded from analyses if they failed these attention checks or did not complete the majority of the survey. This allows researchers to check whether participants are reading each question and exclude those who do not follow the instructions in the “attention check” items. TurkPrime recruited additional participants to replace the excluded subjects. Given a slight overflow in testing, this process yielded our sample of 185 participants.

2.1. Measures

The depression scale was based on a modification of the Clinically Useful Depression Outcome Scale (CUDOS) (Zimmerman et al., 2008). The CUDOS includes 16 items assessing the DSM-IV/DSM-5 symptom criteria for MDD. Compound diagnostic criteria referring to more than one construct (e.g. problems concentrating or making decisions; insomnia or hypersomnia) were subdivided into their respective components and a CUDOS item assesses each component. The individual symptoms assessed by the CUDOS are: depressed mood, loss of interest in usual activities, low energy, psychomotor agitation, psychomotor retardation, guilt, worthlessness, thoughts of death, suicidal ideation, impaired concentration, indecisiveness, decreased appetite, increased appetite, insomnia, hypersomnia, and hopelessness. To the original version of the CUDOS we added 2 items: loss of pleasure in usual activities and low motivation.

The respondent was instructed to rate the symptom items on two 4-point ordinal scales indicating “how well the item describes you during the past week, including today.” Symptom frequency was rated as follows: 1 = 0 days, 2 = 1–2 days, 3 = 3–5 days, 4 = 6–7 days). Symptom severity was rated as 1 = not present, 2 = mild, 3 = moderate, 4 = severe. The 2 sets of symptom ratings were displayed side-by-side.

After completion of the symptom ratings the subjects rated their global level of depression severity. The subjects were asked to rate the “overall level of severity of your symptoms of depression during the past week including today” on a 5-point ordinal scale (none, minimal, mild, moderate, severe).

Psychosocial impairment due to depression and quality of life were assessed with the subscales from the Diagnostic Inventory for Depression (DID) (Zimmerman et al., 2004). The six-item psychosocial functioning subscale assesses the amount of difficulty symptoms of depression have caused in usual daily responsibilities, relationships with significant others such as spouse, relationships with close family members, relationships with friends, participation in leisure activities, and overall level of function. Items are rated on a 5-point ordinal scale (0 = no difficulty; 4 = extreme difficulty). The quality of life subscale assesses satisfaction with the same areas covered by the psychosocial functioning scale as well as global satisfaction with mental health and physical health. Items are rated on a 5-point Likert scale (0 = very satisfied; 4 = very dissatisfied).

2.2. Data analysis

We computed a total score for the intensity ratings, the frequency ratings, and the sum of intensity and frequency ratings. For the 3 subscales (intensity, frequency, sum) we conducted 3 analyses. First, at both the item level and total subscale level, we examined the correlation between the frequency and intensity ratings, and the correlation between each of these subscales with the sum total of frequency and intensity scores. Second, we examined the internal consistency and item-scale correlations for all 3 subscales. And third, we examined the correlations between each subscale and global rating of depression severity, psychosocial functioning, and quality of life. We compared the correlations between these variables with the frequency, intensity, and sum ratings by calculating the difference between the Fisher z transformations of the correlation coefficients and dividing the difference by the standard error (Steiger, 1980).

Table 1
Demographic characteristics of sample (N = 185).

Characteristic	No.	%
Gender		
Male	57	30.8
Female	120	64.9
Transgender	4	2.2
Genderqueer/other	4	2.2
Ethnicity		
White	148	80.0
African-American	21	11.4
Hispanic	2	1.2
Asian	4	2.2
Multiracial	6	3.2
Other	1	0.5
Education		
Less than high school	6	3.2
Graduated high school	122	65.9
Graduated 4 year college or greater	57	30.8
Age (mean ± SD)	39.5 ± 14.7 years	

3. Results

The data in Table 1 shows that the majority of the 185 subjects were female, white, and graduated high school. The mean age of the sample was 39.5 years (SD = 14.7). Most subjects indicated that they were in current treatment for depression (88.7%, n = 165). More patients were taking medication for depression (76.3%, n = 142) than receiving counseling for depression (60.2%, n = 112). Approximately half of the patients were taking medication and receiving counseling (48.6%, n = 90). More than one-third reported a history of suicide attempt (38.2%, n = 71), and nearly half reported having previously being hospitalized for depression (46.2%, n = 86).

The mean total score of the items based on symptom frequency ratings was 43.2 (SD = 12.5) (range: 18–71), similar to the mean total score of the items based on symptom intensity ratings (42.1; SD = 12.3) (range: 18–69). The correlation between the total scale score based on the symptom intensity and symptom frequency ratings was high (r = 0.93, p < .001). Consistent with this, for each symptom the correlation between the symptom frequency and symptom intensity rating was significant (mean = 0.81) (Table 2). We computed a sum of the intensity and frequency scores, and in general the intensity and frequency ratings were equally correlated with sum total (Table 2).

All item scale correlations were significant for the frequency ratings (mean = 0.62), the intensity ratings (mean = 0.64), and the combined frequency/intensity ratings (mean = 0.65) (Table 3). Cronbach's alpha internal consistency coefficients for each scale were nearly identical (frequency, 0.92; intensity, 0.93; sum frequency/intensity, 0.96).

All correlations with the subjects' global ratings of depression severity were significant for the symptom frequency ratings (mean = 0.52), the intensity ratings (mean = 0.51), and the summed frequency/intensity ratings (mean = 0.54) (Table 4). None of the differences between the correlations with the frequency and intensity ratings were significant.

Table 2
Correlation between symptom severity and symptom frequency ratings.

Symptom	Intensity vs. frequency	Intensity vs. sum[a]	Frequency vs. sum[a]
Depressed mood	0.73	0.92	0.94
Lack of interest	0.81	0.95	0.95
Lack of pleasure	0.78	0.94	0.95
Poor appetite	0.82	0.95	0.96
Increased appetite	0.78	0.94	0.95
Low opinion of yourself	0.86	0.96	0.97
Difficulty sleeping	0.76	0.93	0.94
Sleeping too much	0.80	0.95	0.95
Psychomotor physically slowed down	0.81	0.95	0.96
Feeling fidgety or physically restless	0.84	0.96	0.95
Low energy level	0.81	0.95	0.95
Guilty feelings	0.86	0.96	0.97
Problems concentrating	0.81	0.95	0.95
Difficulty making decisions	0.84	0.96	0.96
Wishes to be dead	0.85	0.96	0.96
Suicidal thoughts	0.83	0.96	0.96
Negative view of the future	0.84	0.96	0.96
Low motivation level	0.82	0.95	0.82
Total	0.93	0.98	0.98

All correlations are significant at p < .001.
[a] Sum represents the sum of the items' intensity and frequency ratings

Table 3
Item-scale correlations for symptom intensity and symptom frequency ratings.

Symptom	Symptom intensity	Symptom frequency	Sum[a]
Depressed mood	0.73	0.78	0.80
Lack of interest	0.77	0.76	0.80
Lack of pleasure	0.76	0.73	0.79
Poor appetite	0.49	0.42	0.48
Increased appetite	0.39	0.29	0.35
Low opinion of yourself	0.73	0.76	0.77
Difficulty sleeping	0.51	0.42	0.49
Sleeping too much	0.38	0.31	0.34
Psychomotor physically slowed down	0.71	0.66	0.71
Feeling fidgety or physically restless	0.49	0.43	0.47
Low energy level	0.70	0.76	0.80
Guilty feelings	0.69	0.68	0.71
Problems concentrating	0.74	0.75	0.77
Difficulty making decisions	0.66	0.66	0.67
Wishes to be dead	0.67	0.63	0.67
Suicidal thoughts	0.59	0.53	0.58
Negative view of the future	0.72	0.71	0.74
Low motivation level	0.70	0.72	0.75

All correlations are significant at p < .001.
[a] Sum represents the sum of the items' intensity and frequency ratings.

Table 4
Correlations between symptom severity and symptom frequency ratings and global rating of severity.

Symptom	Symptom intensity	Symptom frequency	Sum[a]
Depressed mood	0.75	0.75	0.80
Lack of interest	0.62	0.68	0.69
Lack of pleasure	0.65	0.63	0.68
Poor appetite	0.39	0.38	0.40
Increased appetite	0.21	0.22	0.23
Low opinion of yourself	0.60	0.66	0.65
Difficulty sleeping	0.47	0.40	0.47
Sleeping too much	0.22	0.22	0.23
Psychomotor physically slowed down	0.50	0.50	0.53
Feeling fidgety or physically restless	0.35	0.32	0.35
Low energy level	0.55	0.57	0.59
Guilty feelings	0.53	0.55	0.56
Problems concentrating	0.56	0.61	0.63
Difficulty making decisions	0.57	0.56	0.58
Wishes to be dead	0.56	0.59	0.60
Suicidal thoughts	0.50	0.49	0.51
Negative view of the future	0.64	0.62	0.66
Low motivation level	0.58	0.64	0.64
Total	0.75	0.79	0.79

All correlations are significant at p < .001, except for increased appetie and sleeping too much (p < .01).
[a] Sum represents the sum of the items' intensity and frequency ratings.

Table 5
Correlation between for symptom severity and symptom frequency total scores and psychosocial functioning and quality of life.

	Symptom intensity	Symptom frequency	Sum[a]
Functioning			
Daily responsibilities (n = 184)	0.73	0.73	0.74
Relationship significant other (n = 127)	0.63	0.61	0.63
Relationship close family (n = 184)	0.65	0.65	0.66
Relationship friends	0.58	0.59	0.60
Leisure activities (n = 184)	0.78	0.76	0.79
Global rating	0.81	0.80	0.82
Completely unable to function	0.61	0.61	0.62
Quality of life			
Daily responsibilities (n = 183)	0.59	0.65	0.63
Relationship significant other (n = 127)	0.35	0.39	0.38
Relationship close family (n = 184)	0.51	0.52	0.53
Relationship friends	0.54	0.55	0.56
Leisure activities	0.62	0.62	0.63
Mental health	0.65	0.70	0.69
Physical health	0.55	0.59	0.58
Global rating of satisfaction	0.61	0.63	0.63
Overall quality of life	0.62	0.67	0.66

All correlations are significant at $p < .001$.
[a] Sum represents the sum of the items' intensity and frequency ratings.

Similarly, all correlations with the ratings of functional impairment and quality of life were significant and nearly identical for the total score based on frequency ratings (mean = 0.63), intensity ratings (mean = 0.61), and the combined frequency/intensity ratings (mean = 0.63) (Table 5). None of the differences in the correlations with the frequency and intensity ratings were significant.

4. Discussion

Illness severity is a clinically important construct. While the research has not been entirely consistent, the severity of depression has been associated with health-related quality of life (Fattori et al., 2017), functional impairment (Goethe et al., 1993; Luty et al., 2002), suicidality (Bradvik et al., 2008; Kessing, 2004; Wang et al., 2015), longitudinal course (Katon et al., 2010; Keller et al., 1992; Melartin et al., 2004; Meyers et al., 2002), and several biological variables (Berent et al., 2014; de Diego-Adelino et al., 2014; Zimmerman et al., 1986). Moreover, the severity of depression has been at the core of controversies regarding the efficacy of treatment and whether certain forms of treatment should be recommended as first line interventions (Zimmerman, 2019). Almost all research on severity is based on scores on depression symptom scales. Most scales have been developed without consideration as to how to best conceptualize and assess the severity of depression.

In the present study we found that symptom ratings based on intensity and frequency were highly correlated with each other, and equally highly correlated with subjects' global rating of overall severity of depression, as well as ratings of psychosocial functioning and quality of life. Moreover, a composite index of severity based on the sum of frequency and severity ratings was no more highly correlated with the external validators. Thus, the results of the present study suggest that ratings based on either symptom intensity or symptom frequency are equally valid.

There are, however, circumstances in which symptom intensity ratings would be the preferred metric. In studies of rapidly effective treatments in which outcome is assessed on a daily basis, it does not make sense to assess the frequency of some symptoms such as sleep and appetite disturbance. The instructional set for symptom ratings thus needs to be appropriate to the assessment interval, and when brief intervals are used to rate the items, then symptom intensity ratings are more appropriate (Zimmerman et al., 2018).

It is also possible that in certain subgroups of patients that intensity and severity ratings will not be equally associated with external validators. For example, perhaps in a sample of patients with bipolar disorder, who experience mood swings, the correlations between validators with duration ratings will be less than the correlations with intensity ratings.

Our sample was obtained through an online recruitment platform. Online recruitment provides a number of advantages over traditional data collection methods including greater affordability, efficiency, ease in collecting large samples of data, and ability to recruit samples that are more representative and diverse than typical subject pools. Prime panels are more diverse than MTurk and more closely resemble the US population. Though our sample consists of majority white, high school educated females, this reflects the characteristics of the particular participant pool that reported receiving treatment for a mental health problem and is likely more representative than typical research samples obtained at one clinical practice or in a restricted geographical area.

Online crowdsourcing platforms such as MTurk or TurkPrime usually provide valid and high-quality data (Buhrmester et al., 2011; Litman et al., 2015; Miller et al., 2017; Shapiro et al., 2013). However, methods such as MTurk sometimes produce poor quality data due to participant familiarity with similar study procedures, dishonesty, or inattention (Chandler and Paolacci, 2017). The TurkPrime prescreening method reduces the issue of dishonesty by asking participants if they are in mental health treatment through ongoing behavioral questions without compensation rather than to determine inclusion criteria for a study (Chandler and Shapiro, 2016; Chandler and Paolacci, 2017). To address the possible issue of inattention, we identified and removed poor quality data by excluding participants who failed our attention-check questions. However, online crowdsourcing is unable to control the testing environment or verify responses, and some researchers argue that these check questions do not sufficiently measure attention (Hauser and Schwarz, 2015).

An additional limitation of the present study and of online crowdsourcing in general is the reliance on self-report measures. In the present case, the applicability to clinician-based ratings of symptomatology is uncertain.

We examined validity from one perspective—correlation with measures functioning and quality of life. Other approaches to comparing the validity of measures of depression severity include prediction of outcome, distinguishing patients who do and do not require hospitalization, correlation with biological parameters of depression, and demonstration of a larger effect size in treatment studies.

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2019.112512.

References

Berent, D., Zboralski, K., Orzechowska, A., Galecki, P., 2014. Thyroid hormones association with depression severity and clinical outcome in patients with major depressive disorder. Mol. Biol. Rep. 41, 2419–2425.
Bradvik, L., Mattisson, C., Bogren, M., Nettelbladt, P., 2008. Long-term suicide risk of depression in the Lundby cohort 1947-1997–severity and gender. Acta Psychiatr. Scand. 117, 185–191.
Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6, 3–5.
Chandler, J., Shapiro, D., 2016. Conducting clinical research using crowdsourced convenience samples. Annu. Rev. Clin. Psychol 12, 53–81.

Chandler, J.J., Paolacci, G., 2017. Lie for a dime: when most prescreening responses are honest but most study participants are impostors. Soc. Psychol. Personal. Sci. 8, 500–508.
Cheung, H.N., Power, M.J., 2012. The development of a new multidimensional depression assessment scale: preliminary results. Clin. Psychol. Psychother. 19, 170–178.
Davidai, S., 2018. Why do Americans believe in economic mobility? Economic inequality, external attributions of wealth and poverty, and the belief in economic mobility. J. Exp. Soc. Psychol 79, 138–148.
de Diego-Adelino, J., Pires, P., Gomez-Anson, B., Serra-Blasco, M., 2014. Microstructural white-matter abnormalities associated with treatment resistance, severity and duration of illness in major depression. Psychol. Med. 44, 1171–1182.
Fattori, A., Neri, L., Bellomo, A., Vaggi, M., 2017. Depression severity and concentration difficulties are independently associated with HRQOL in patients with unipolar depressive disorders. Qual. Life Res on line May 13, 2017.
Goethe, J.W., Fischer, E.H., Wright, J.S., 1993. Severity as a key construct in depression. J. Nerv. Ment. Dis. 181, 718–724.
Hauser, D.J., Schwarz, N., 2015. It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. Sage Open 5, 2158244015584617.
Katon, W., Unutzer, J., Russo, J., 2010. Major depression: the importance of clinical characteristics and treatment response to prognosis. Depress. Anxiety 27, 19–26.
Keller, M.B., Lavori, P.W., Mueller, T.I., Endicott, J., 1992. Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. Arch. Gen. Psychiatry 49, 809–816.
Kessing, L.V., 2004. Severity of depressive episodes according to ICD-10: prediction of risk of relapse and suicide. Br. J. Psychiatry 184, 153–156.
Kroenke, K., Spitzer, R., Williams, J., 2001. The PHQ-9. Validity of a brief depression severity measure. J. Gen. Int. Med. 16, 606–613.
Litman, L., Robinson, J., Rosenzweig, C., 2015. The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on mechanical Turk. Behav. Res. Methods 47, 519–528.
Luty, S.E., Joyce, P.R., Mulder, R.T., McKenzie, J.M., 2002. Social adjustment in depression: the impact of depression severity, personality, and clinic versus community sampling. J. Affect. Disord. 70, 143–154.
Melartin, T., Rytsala, H., Leskela, U., Lestela-Mielonen, P., 2004. Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder. J. Clin. Psychiatry 65, 810–819.
Meyers, B.S., Sirey, J.A., Bruce, M., Hamilton, M., 2002. Predictors of early recovery from major depression among persons admitted to community-based clinics: an observational study. Arch. Gen. Psychiatry 59, 729–735.
Miller, J.D., Crowe, M., Weiss, B., Maples-Keller, J.L., 2017. Using online, crowdsourcing platforms for data collection in personality disorder research: the example of Amazon's mechanical Turk. Pers. Disord. 8, 26–34.
Olsen, L.R., Jensen, D.V., Noerholm, V., Martiny, K., 2003. The internal and external validity of the major depression inventory in measuring severity of depressive states. Psychol. Med. 33, 351–356.
Parker, G., Hadzi-Pavlovic, D., Sengoz, A., Boyce, P., 1994. A brief self-report depression measure assessing mood state and social impairment. J. Affect. Disord. 30, 133–142.
Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., 2011. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS(R)): depression, anxiety, and anger. Assessment 18, 263–283.
Rush, A., Trivedi, M., Ibrahim, H., Carmody, T., 2003. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psychiatry 54, 573–583.
Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., 1996. The inventory of depressive symptomatology (IDS). Psychol. Med. 26, 477–486.
Shapiro, D.N., Chandler, J., Mueller, P.A., 2013. Using mechanical Turk to study clinical populations. Clin. Psychol. Sci. 1, 213–220.
Skrzynski, C., Creswell, K.G., Bachrach, R.L., Chung, T., 2018. Social discomfort moderates the relationship between drinking in response to negative affect and solitary drinking in underage drinkers. Addict. Behav. 78, 124–130.
Steiger, J.H., 1980. Tests for comparing elements of a correlation matrix. Psychol. Bull. 87, 245–251.
Wang, Y.Y., Jiang, N.Z., Cheung, E.F., Sun, H.W., 2015. Role of depression severity and impulsivity in the relationship between hopelessness and suicidal ideation in patients with major depressive disorder. J. Affect. Disord. 183, 83–89.
Williams, J.B., Kobak, K.A., Bech, P., Engelhardt, N., 2008. The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. Int. Clin. Psychopharmacol. 23, 120–129.
Zimmerman, M., 2019. Severity and the treatment of depression: a review of two controversies. J. Nerv. Ment. Dis. 207, 219–223.
Zimmerman, M., Chelminski, I., McGlinchey, J.B., Posternak, M.A., 2008. A clinically useful depression outcome scale. Compr. Psychiatry 49, 131–140.
Zimmerman, M., Coryell, W., Pfohl, B., 1986. The validity of the dexamethasone suppression test as a marker for endogenous depression. Arch. Gen. Psychiatry 43, 347–355.
Zimmerman, M., Galione, J., Attiullah, N., Friedman, M., 2011. Depressed patients perspectives of two measures of outcome: the Quick Inventory of Depressive Symptomatology (QIDS) and the Remission from Depression Questionnaire (RDQ). Ann. Clin. Psychiatry 23, 208–212.
Zimmerman, M., Harris, L., Martin, J., McGonigal, P., 2018. Reliability and validity of a self-report scale for daily assessments of the severity of depressive symptoms. Psychiatry Res. 270, 581–586.
Zimmerman, M., Sheeran, T., Young, D., 2004. The diagnostic inventory for depression: a self-report scale to diagnose DSM-IV for major depressive disorder. J. Clin. Psychol. 60, 87–110.