Research paper

# Assessing depression improvement with the remission evaluation and mood inventory tool (REMIT)

Michael A. Bushey[a,*], Kurt Kroenke[b,c,d], Fitsum Baye[e], Spencer Lourens[e]

[a] Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States
[b] VA HSR&D Center for Health Information and Communication, Roudebush VA Medical Center, Indianapolis, IN, United States
[c] Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States
[d] Regenstrief Institute, Inc., Indianapolis, IN, United States
[e] Department of Biostatistics, Indiana University Fairbanks School of Public Health, Indianapolis, IN, United States

ARTICLE INFO

ABSTRACT

*Objective:* The Remission Evaluation and Mood Inventory Tool (REMIT) was developed as a brief complementary measure to provide a more robust assessment of depression improvement than tracking DSM-V symptom improvement alone. This study provides further validation of the REMIT tool and examines its utility in predicting depression improvement.

*Methods:* The sample comprised 294 primary care patients enrolled in a telecare trial of pain plus depression and/or anxiety. Assessments collected included: REMIT, PHQ-9 and measures assessing anxiety, pain, sleep, fatigue, somatization, health-related quality of life and disability. Data was analyzed to assess the REMIT's validity, its minimally important difference (MID), and its utility in predicting 6-month depression improvement.

*Results:* Convergent and construct validity of REMIT was supported by moderate correlations with mental health measures and weaker correlation with physical health measures. MID of approximately 2 points for REMIT was estimated by two metrics: 0.5 standard deviation and 1 standard error of measurement. Both baseline and 3-month change in REMIT scores predicted depression improvement at 6 months. Indeed, REMIT was as good or better predictor than the PHQ-9.

*Conclusion:* The REMIT measure is a brief 5-item tool that augments core DSM-V symptom-oriented metrics in assessing and predicting recovery from major depression.

## 1. Introduction

Depression is a common clinical problem, with a prevalence of up to 20% in the community [1]. It is the most common mental disorder in both primary care and mental health specialty settings and the leading cause of disability among all diseases [2,3]. Despite the clinical and public health impact of depression, the guidelines for what constitutes successful treatment can be difficult to define. Many have advocated a "treat to remission" approach, wherein clinicians adjust treatment until patients reach certain thresholds, often defined as scoring 7 or less on the Hamilton Rating Scale for Depression (HRSD-17) or below 5 on the Patient Health Questionnaire-9 (PHQ-9) [4,5]. Unfortunately, this approach can lead to both under and over-treatment. Since all measures have imperfect sensitivity and specificity, false positives and false negatives are expected for any cut-point chosen. Unlike patients in clinical trials, real-world patients often have significant medical comorbidities which can cause impairments in sleep, appetite, energy, and

concentration, resulting in persistent elevation in these assessment domains. On the other hand, Zimmerman et al. have found that a significant proportion of patients who meet criteria for remission as defined by HRSD-17 scores <7 do not consider themselves in remission [6]. This suggests that there may be aspects of recovery from illness not captured by assessing only the core features of DSM-defined major depressive disorder.

These observations led to the development of the Remission Evaluation and Mood Inventory Tool (REMIT), a 5-item self-administered questionnaire which measures aspects of depression recovery not captured using core-symptom focused tools such as the PHQ-9 or HRSD-17 [7]. The REMIT items are more focused on positive psychology constructs such as happiness, contentment, emotional regulation, resilience, and hope. This tool performed well in a cross-sectional study, with clearly distinct psychometric properties from the PHQ-8, demonstrating promise as a tool that might complement existing measures to give a more robust account of depression recovery [7]. To date, the

---

REMIT has been studied in two samples. One of these was a cross-sectional study [7,8] and the other had very short-term longitudinal follow-up of only 4 weeks [9]. Hence, the value of REMIT in predicting depression remission or providing a more robust account of depression recovery over more clinically-relevant time spans has not yet been investigated.

In this study, we postulate that the REMIT tool may complement a measure of DSM-V symptoms (the PHQ-9) in predicting improvement in major depression and patient-reported recovery. We hypothesized that patients with lower REMIT scores at baseline and greater reduction in REMIT scores by 3 months might be more likely to experience depression improvement by 6 months compared to patients with higher baseline REMIT scores and less change at 3 months.

## 2. Methods

### 2.1. Study participants

This study utilized data collected for the CAMMPs trial, detailed previously [10]. Briefly, participants were recruited from five primary care clinics at a large VA Medical Center, and randomized to two levels of telecare intervention for pain and mood symptoms in a 12-month pragmatic trial. The less intensive treatment arm received automated symptom monitoring plus web-based self-management, while the more intensive arm received the same interventions plus telephone-based treatment monitoring and adjustment by a nurse care manager-physician specialist team. Eligibility criteria included age ≥ 18 years old and clinically significant musculoskeletal pain plus psychiatric comorbidity. Pain had to be moderate in severity as assessed by the brief pain inventory (BPI) and present at least 3 months despite analgesic treatment. Psychiatric comorbidity was defined as depression or anxiety of at least moderate severity with either: [1] a score of 10 or greater on the PHQ-8 with anhedonia and/or depressed mood endorsed; [2] a score of 10 or greater on the GAD-7; or [3] a combined PHQ-8 and GAD-7 score of at least 12. The PHQ-8 was used for eligibility determination since it omits the item about self-harm yet has scoring thresholds identical to the PHQ-9 [11,12]. Exclusion criteria included: not speaking English; moderately severe cognitive impairment; suicidality, schizophrenia, bipolar disorder, or another severe or complex mental illness not suitable for telecare; being pregnant; or an anticipated life expectancy of < 12 months.

### 2.2. Study measures

#### 2.2.1. Measure administration

Measures to assess depression, anxiety, pain, health-related quality of life, disability, medical co-morbidity, and global impression of change were collected at baseline and at defined time points during the CAMMPS trial [10]. All measures were administered by a research interviewer: in person at baseline and over the phone for follow-up time points.

#### 2.2.2. Depression measures

The Patient Health Questionnaire 9-item depression scale (PHQ-9) is a self-administered 9-item questionnaire wherein patients rate whether they have experienced each of the described symptoms "not at all", "several days", "more than half the days" or "nearly every day" over the preceding two weeks [13]. Scores per item range from 0 to 3, with higher scores indicating more severe symptoms, and a total score range of 0–27. Using the PHQ-9, a patient would meet DSM-V criteria for major depression (MD) if they endorsed at least 5 of 9 items as present for more than half of the past 14 days and one item must be depressed mood or anhedonia. This diagnostic algorithm was used to determine the presence or absence of MD in this study.

The Remission Evaluation and Mood Inventory Tool (REMIT) is a 5-item self-administered questionnaire that was developed to measure

aspects of recovery from depression not captured using core-symptom focused tools such as the PHQ-9 [7]. Patients are asked to rate whether - over the past two weeks - they felt happy, content, in control of emotions, bounced back when things went wrong or whether the future seemed dark (the latter item is reverse-scored). Choices include "none of the time", "a little of the time", "some of the time", "most of the time" and "all of the time". The scores for each item range from 0 to 4, with a possible total score range of 0–20 and higher scores reflecting worse severity.

#### 2.2.3. Other measures

The Generalized Anxiety Disorder Scale (GAD-7) is a 7-item self-administered questionnaire that assesses the severity of anxiety experienced over the past two weeks. Scores range from 0 to 21, with higher scores reflecting greater levels of anxiety [14].

The Patient Reported Outcomes Measurement Information System (PROMIS) includes a variety of measures to evaluate physical and mental health [15,16]. The 8-item scales for anxiety and depression from the PROMIS-57 profile were collected, along with the 4-item scales for assessing sleep and fatigue from the PROMIS-29 profile. Raw scores are standardized to a T-score for which the general population mean is 50 with standard deviation of 10; higher scores reflect more severe symptoms.

The Brief Pain Inventory (BPI) is an 11-item self-administered instrument that assesses pain severity (4 items) and pain interference (7 items) [17]. Each item is scored from 0 to 10 and total BPI score is the mean of the 11 items with higher scores representing worse pain.

The Patient Health Questionnaire somatic symptom scale (PHQ-15) assesses the severity of 15 common somatic symptoms. The PHQ-14 is a modified version of this scale in which the sexual dysfunction question is omitted [18]. PHQ-14 scores range from 0 to 28, with higher scores reflecting greater somatic symptom severity.

Health-related quality of life was assessed using the Medical Outcomes Study Short-Form 12-item scale (SF-12) from which the Mental Component Summary (MCS) and Physical Component Summary (PCS) scores can be derived [19]. In addition to the SF-12, 8 items were included from the SF-36, which allowed assessment of the SF-36 social functioning, vitality, bodily pain, mental health, and current general health scores [20]. Scores on these scales range from 0 to 100 with lower scores representing worse health-related quality of life.

Three disability measures were administered. The Sheehan Disability Scale (SDS) asks how much health has interfered with family life, social life, and work over the past month on a scale of 0 (*not at all*) to 10 (*unable to carry on any activities*) [21]. The SDS score is a mean of these three items with higher scores reflecting greater disability. Disability days were assessed with a single item that asked participants to indicate the number of days during the preceding 4 weeks that they were either in bed or had to reduce work or usual activities by 50% or more due to physical health or emotional problems [10]. Work effectiveness was assessed with a single item asking how effective the respondent was on his or her job during the past 4 weeks on a scale of 0% (not at all effective) to 100% (completely effective).

Medical co-morbidity was assessed using a checklist that asks about the presence or absence of 9 common medical disorders [22].

Global impression of change was a single item asking patients to rate mood or anxiety symptoms compared to the start of the trial as "much worse", "moderately worse", "a little worse", "same", "a little better", "moderately better", or "much better" [10].

### 2.3. Statistical analysis

SAS Version 9.3 was used for all analyses. Correlations between REMIT and PHQ-9 individual items as well as total scores were calculated. Convergent validity was assessed by calculating correlations of the REMIT and PHQ-9 with mental health measures including the SF-12 MCS, SF-36 mental health, PROMIS depression and anxiety, and the

**Table 1**
Pearson correlations between REMIT and PHQ-9 ($n = 294$).

| REMIT items | PHQ-9 items | | | | | | | | | PHQ-9 total score |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 anhedonia | 2 depressed mood | 3 sleep | 4 appetite | 5 fatigue | 6 guilt | 7 concentration | 8 psychomotor | 9 self-harm | |
| Did you feel happy? | 0.31 | 0.40 | 0.11 | 0.25 | 0.01 | 0.32 | 0.15 | 0.08 | 0.26 | 0.39 |
| Did you feel content? | 0.36 | 0.40 | 0.01 | 0.26 | 0.11 | 0.34 | 0.20 | 0.05 | 0.23 | 0.39 |
| Did you feel in control of your emotions? | 0.18 | 0.19 | 0.11 | 0.21 | 0.09 | 0.33 | 0.21 | 0.13 | 0.07 | 0.32 |
| Did you bounce back when things went wrong? | 0.15 | 0.28 | 0.05 | 0.16 | 0.00 | 0.28 | 0.18 | 0.15 | 0.18 | 0.30 |
| Did the future seem dark to you? | 0.30 | 0.34 | −0.04 | 0.17 | −0.05 | 0.36 | 0.19 | 0.14 | 0.32 | 0.37 |
| REMIT total score | 0.36 | 0.45 | 0.06 | 0.29 | 0.04 | 0.46 | 0.26 | 0.16 | 0.30 | 0.50 |

GAD-7. Construct validity was assessed by calculating correlations with the Sheehan Disability Index, disability days, work effectiveness, PHQ-14 somatization, PROMIS sleep and fatigue, and SF vitality, social functioning, and general health.

To examine the predictive value of REMIT, we conducted three types of analyses.

1) We modeled whether baseline scores predicted major depression (MD) status at 6 months in the subgroup of 135 patients with MD at baseline. Logistic regression models were run with MD status at 6 months (yes or no) as the dependent variable and baseline REMIT and PHQ-9 scores entered as predictors individually and then together in the same model. The association between other patient characteristics and MD status at 6 months was analyzed to determine if models needed to be adjusted for other covariates. The Akaike Information Criterion (AIC) was calculated for each model to determine the model with the best fit (i.e., lowest AIC value).
2) We examined whether change in REMIT scores from baseline to 3 months predicted MD status at 6 months and, as a sensitivity analysis, three other ways of assessing depression improvement: global improvement in mood, PHQ-9 response (score <10) and PHQ-9 remission (score <5). Here, we used receiver operating curve analysis to compare the areas under the curve (AUC) for the four different ways of assessing depression improvement.
3) We used repeated measures logistic regression modeling to determine whether change in REMIT, PHQ-9, and PHQREMIT (sum of REMIT and PHQ-9) scores predicted global improvement in mood over 6 months. Here, we used baseline to 3-month score change to predict 3-month global improvement and 3 to 6-month score change to predict 6-month global improvement. For this analysis, the full sample of 250 patients with follow-up assessments was included for three reasons: 1) all patients in the sample had either depression, anxiety, or both; 2) the global question asks about change in "low mood or anxiety"; 3) the global outcome question is not based upon DSM-V criteria (as is MD status) and thus would be less likely to be biased in favor of the PHQ-9. In contrast to the first two analyses which focused on 6-month major depression status and for which the appropriate subgroup was those with MD at baseline, the third analysis examined improvement in depressive or anxiety symptoms and thus could take advantage of the full sample in assessing the predictive value of REMIT.

## 3. Results

### 3.1. Study participants

A total of 294 participants were enrolled in the study. The overall sample had a mean age of 57.4 years (range 25 to 88), and was 87.4% male, 79.3% white, 15.3% black, and 5.4% other race. Regarding education, 21% were college graduates, 53% had completed some college or trade school, and 26% had only a high school diploma or

GED. Major depression was present at baseline in 135 participants as determined using the PHQ-9 diagnostic algorithm. For purposes of this psychometric analysis, participants from both arms of the trial were collapsed into a single dataset.

### 3.2. REMIT psychometric characteristics

REMIT score at baseline for the 294 study participants was a mean (SD) of 9.0 (3.8). Baseline scores were normally distributed (Fig. S1). Internal reliability was reasonable with a Cronbach's alpha of 0.76, compared to 0.86 and 0.82 in the previous studies [7,9]. The standard error of measurement (SEM), calculated as SD x square root of (1 − Cronbach's alpha), was 1.86; one SEM is typically considered the lower statistical bound of a minimally important difference (MID). Another estimate of MID is 0.5 SD, which was 1.9 in our sample. Taken together, these findings suggest that a 2-point change may represent a reasonable MID approximation for the REMIT scale.

### 3.3. Comparison of REMIT and PHQ-9

REMIT and PHQ-9 total scores were only moderately correlated ($r = 0.50$), similar to what had been found in a previous study comparing the REMIT and PHQ-8 [8]. This substantiates the complementary rather than redundant nature of the REMIT and PHQ-9 scales which focus on positive and negative aspects of mood, respectively. Consistent with the overall scale findings, individual REMIT and PHQ-9 item correlations also ranged from low to moderate as shown in Table 1.

Of the 45 item-item correlations, the median was 0.18, with 11 being 0.30–0.40, 9 being 0.20–0.29, 15 being 0.10–0.19, and 10 being <0.10. Correlations tended to be lower between REMIT items and PHQ-9 somatic items such as sleep, appetite, fatigue, and psychomotor activity, and tended to be higher when compared to psychological items such as anhedonia, mood, or guilt. Spearman correlations were also calculated and were very similar to the Pearson correlations, with the exception being that correlations of the 5 REMIT items with the PHQ-9 fatigue item tended to be higher with the Spearman correlation (median = 0.14, range 0.09 to 0.15) compared to the Pearson correlation (median = 0.01, range 0.05 to 0.11), suggesting that fatigue responses may be more skewed than the other items.

### 3.4. Convergent and construct validity

REMIT and PHQ-9 had relatively similar correlations with the other measures of convergent and construct validity (Table 2), although the PHQ-9 tended to correlate more strongly with measures that included physical symptoms (PHQ-14, PROMIS fatigue and sleep, SF-36 vitality) and measures of disability. The REMIT tended to be more highly correlated with several psychological measures such as SF-12 MCS and SF-36 mental health. As expected, both REMIT and PHQ-9 correlated poorly with medical comorbidity.

**Table 2**

REMIT and PHQ-9 Pearson correlations with other measures ($n = 294$)[a].

| Measure | REMIT | PHQ-9 |
|---|---|---|
| Convergent validity | | |
| Mental health (SF-36) | − 0.76 | − 0.59 |
| Depression (PROMIS 8-item) | 0.70 | 0.61 |
| Mental component summary (SF-12) | − 0.64 | − 0.57 |
| Anxiety (PROMIS 8-item) | 0.53 | 0.48 |
| Anxiety (GAD-7) | 0.46 | 0.52 |
| | | |
| Construct validity | | |
| Vitality (SF-36) | − 0.47 | − 0.59 |
| Social functioning (SF-36) | − 0.45 | − 0.54 |
| Sheehan disability index | 0.41 | 0.54 |
| General health (SF-36) | − 0.34 | − 0.31 |
| Work effectiveness, percent | − 0.31 | − 0.38 |
| Disability days in past 4 weeks | 0.29 | 0.41 |
| Fatigue (PROMIS 4-item) | 0.29 | 0.55 |
| Somatization (PHQ-14) | 0.27 | 0.51 |
| Sleep (PROMIS 4-item) | 0.22 | 0.37 |
| Medical comorbidity (number of diseases) | 0.14 | 0.12 |

[a] Negative correlations are with measures where higher score is better.

### 3.5. Baseline REMIT and PHQ-9 scores in predicting depression improvement

Of the 135 patients who had major depression (MD) at baseline, 70 (52%) no longer met criteria for MD at 6 months, and 65 (48%) still had MD. Baseline characteristics were similar between these two groups (Table S1) except that baseline BPI pain scores were higher in patients who continued to meet MD criteria at 6 months (6.6 vs. 5.9). Logistic regression models were used to examine predictors of depression improvement defined as no longer meeting criteria for major depression at 6 months. When entered individually into logistic regression models, higher (worse) REMIT, PHQ-9, and BPI pain baseline scores each predicted a lower odds of depression improvement. When entered in pairwise combinations, REMIT and BPI pain (but not PHQ-9) scores remained predictive. When all 3 scores were entered into the same model, BPI remained significant ($P = .047$), REMIT was borderline ($P = .07$), and PHQ-9 was not significant ($P = .48$). Using the lowest AIC as a criterion, the pairwise model including REMIT and BPI pain appeared optimal (Table S2).

### 3.6. Change in REMIT score in predicting 4 ways of measuring depression improvement

To assure that our results were not influenced by the particular method for determining depression improvement, we compared 4 different ways of characterizing 6-month depression improvement: major depression status (i.e., no longer meeting criteria for MD), global improvement, and PHQ-9 response and remission (scores < 10 and 5, respectively). The AUCs for REMIT change were similar across all 4 types of depression outcome (Table 3).

**Table 3**

REMIT change at 3-months as a predictor of four ways of defining 6-month depression improvement in 135 patients with major depression at baseline.

| Depression improvement outcome at 6 months | Frequency | Percent | AUC for REMIT change at 3 months | |
|---|---|---|---|---|
| | | | AUC | SE |
| No major depression | 70 | 51.9 | 0.592 | 0.049 |
| Global improvement in mood | 54 | 40.0 | 0.595 | 0.049 |
| PHQ-9 < 10 (response) | 34 | 25.2 | 0.554 | 0.056 |
| PHQ-9 < 5 (remission) | 16 | 11.9 | 0.563 | 0.066 |

### 3.7. Change in scale scores predicting global improvement in mood

In addition to baseline scores as predictors, we hypothesized that early change in scores might also be predictive of depression improvement. Table 4 summarizes results of the repeated measures logistic regression modeling examining change in REMIT and PHQ-9 scores as well as the composite PHQREMIT score (sum of REMIT and PHQ-9) in predicting global improvement in mood over 6 months. Improvement in each of the scores was predictive of global improvement. Although odds ratios were relatively similar, REMIT improvement had the highest OR and lowest AIC values in models with REMIT alone or with the addition of PHQ-9.

### 4. Discussion

Our study has several important findings. First, the REMIT scale has good internal reliability as well as convergent and construct validity. Second, a 2-point change in the REMIT score appears to be a reasonable estimate of the minimally important difference. Our 1-SEM and 0.5 SD estimates of MID were both 1.9, which correspond to estimates calculated from data in Nease et al. (1.7 and 2.3) and Zeng et al. (1.2 and 1.4) [7,9]. Third, the REMIT score compared to the PHQ-9 score was a stronger predictor of 6-month depression outcomes whether using baseline or change scores.

The REMIT has previously been shown to provide a more robust picture of improvement from major depression than tracking DSM symptoms alone. It was originally validated using a cross-sectional sample of 1003 primary care patients who were administered the PHQ-8 and REMIT. In that study, a moderate correlation ($r = 0.46$) was found between the PHQ-8 and REMIT, similar to the correlation of the PHQ-9 and REMIT ($r = 0.50$) in our study. REMIT was also examined in 9855 depressed outpatients in China and compared to Hamilton Rating Scale for Depression (HRSD-17) scores at baseline and after 2 and 4 weeks of antidepressant treatment. REMIT and HRSD-17 scores were moderately correlated at baseline ($r = 0.56$), and both scores improved with treatment. Although the authors suggested that REMIT scores lagged behind HRSD-17 change, the effect size improvements (i.e., change divided by baseline SD) in HSRD-17 and REMIT scores were relatively similar at both 2 weeks (1.69 and 1.46, respectively) and 4 weeks (2.76 and 2.67). Moreover, the 2- and 4-week changes in REMIT score of 4.1 and 7.5 points were large given the 2-point MID estimate. Strengths of our study include longer follow-up (6 months compared to the previous cross-sectional and 4-week studies), a rich battery of measures to assess convergent and construct validity, and several methods for assessing depression outcome.

Treating depression to remission in measurement-based care entails adjusting therapy until patients achieve a desired outcome. Score cut-points on metrics of depression, such as below 5 on PHQ-9 or below 7 on HRSD-17 are often used as benchmarks of remission. Patients and clinicians have noted that such definitions may not fully capture the patient experience, as a number of patients with metric-defined remission do not perceive themselves as well [6]. This feeling of wellness is better captured in the concept of recovery, which can be defined to encompass an improvement of not only the core symptoms of an illness but also quality-of-life and subjective well-being. This is in contrast to defining recovery as sustained remission over time [23], which has not proven optimal in predicting sustained improvement or decreased risk of relapse [24]. REMIT was developed to assess additional facets of depression recovery to provide a more robust account of improvement. Our data is consistent with prior samples demonstrating that administration of the REMIT as an adjunct to symptom tracking with PHQ-9 is a more powerful predictor of depression improvement than PHQ-9 alone. This is consistent whether looking at PHQ-9-defined remission or recovery-oriented measures such as global improvement ratings. Although other tools have been developed for this purpose [25], the brevity of the 5-question REMIT is an advantage in busy clinic settings.

**Table 4**
Improvement in scale scores as predictors of global improvement in mood at 6 months in the full sample of participants ($n = 250$).

| Model | Predictor (score change)[a] | Beta (SE) | Odds ratio[b] | 95% CI | | P-value | AIC |
|---|---|---|---|---|---|---|---|
| | | | | Outcome: 6-month global mood improvement | | | |
| Model 1a | REMIT improvement | 0.19 (0.04) | 1.22 | 1.12 | 1.31 | <0.0001 | 626.23 |
| Model 1b | PHQ-9 improvement | 0.08 (0.02) | 1.09 | 1.04 | 1.14 | 0.0008 | 642.45 |
| Model 1c | PHQREMIT improvement | 0.10 (0.02) | 1.10 | 1.06 | 1.14 | <0.0001 | 626.08 |
| Model 2 | REMIT improvement | 0.17 (0.04) | 1.19 | 1.10 | 1.29 | <0.0001 | 622.77 |
| | PHQ-9 improvement | 0.06 (0.03) | 1.06 | 1.01 | 1.12 | 0.02 | |

[a] Repeated measures logistic regression modeling using baseline to 3 month score change to predict 3-month global improvement and 3 to 6-month score change to predict 6-month global improvement.

[b] An odds ratio >1.0 means that reduction in the scale score increases the odds of improvement. For example, an OR of 1.19 for REMIT means that every 1-point reduction in REMIT score increases the odds of global improvement at 6 months by 19%.

Our study sample was limited by its relative homogeneity. Patients were enrolled from a primary care population of predominantly male veterans who all had comorbid chronic pain. Future studies in non-veteran samples with more females, in patients without chronic pain, and in individuals cared for in mental health settings will be required to better estimate generalizability. Additionally, the AUCs in Table 4 represent modest rather than strong differentiation between patients whose depression had improved and those who were the same or worse. However, AUCs have fallen in a similar range in other studies using patient-rated global improvement (PGIC) as an anchor [26–29] in which AUCs tend to be lower than in studies of diagnostic tests for which there is a criterion ("gold") standard to determine the presence of a disease. PGIC requires patients to retrospectively assess change in symptoms compared to a prior time point and may be influenced by recall bias as well as the current state of symptoms. Thus, the similarity of AUCs when using a PGIC anchor is more salient than their absolute value and suggests relative comparability of REMIT change for detecting depression improvement measured four different ways.

Although developed to provide a fuller picture of recovery from major depression, the domains assessed by REMIT may also be relevant across a range of mental health conditions including anxiety, substance use, and personality disorders, among others. REMIT is comprised of positive psychology items, which are cross-cutting and therefore warrant further research in both depressive and non-depressive disorders. Several tools for assessing overall well-being have been developed since positive psychology was popularized by Seligman and Csikszentmihalyi [30,31]. Originally designed to complement the PHQ-9, REMIT has the advantage of adding assessment depth without adding redundancy in settings already utilizing PHQ-9 for tracking depression. However, REMIT is not intended to provide a comprehensive assessment of overall well-being or flourishing. Tools that assess multiple domains of well-being, including meaning, purpose, character, social relationships, and financial stability are better suited for that purpose [32]. Given its promise, further studies are warranted examining the sensitivity to change of REMIT in treatment trials and its added value when used in conjunction with standard depression measures.

## Funding

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://

doi.org/10.1016/j.genhosppsych.2019.07.007.

## References

[1] Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003;289(23):3095–105.

[2] Kroenke K, Unutzer J. Closing the false divide: sustainable approaches to integrating mental health services into primary care. J Gen Intern Med 2017;32:404–10.

[3] Olfson M, Kroenke K, Wang S, Blanco C. Trends in office-based mental health care provided by psychiatrists and primary care physicians. J Clin Psychiatry 2014;75:247–53.

[4] Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6(4):278–96.

[5] 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;282:1737–44.

[6] Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. Am J Psychiatry 2006;163(1):148–50.

[7] Nease Jr. DE, Aikens JE, Klinkman MS, Kroenke K, Sen A. Toward a more comprehensive assessment of depression remission: the Remission Evaluation and Mood Inventory Tool (REMIT). Gen Hosp Psychiatry 2011;33(3):279–86.

[8] Aikens JE, Klinkman MS, Sen A, Nease DE. Improving the assessment of depression remission with the Remission Evaluation and Mood Inventory Tool. Int J Psychiatry Med 2015;50(4):383–97.

[9] Zeng Q, Wang WC, Fang Y, Mellor D, McCabe M, Byrne L, et al. Evaluations of treatment efficacy of depression from perspective of both patients' symptoms and general sense of mental health and wellbeing: a large scale, multi-centered, longitudinal study in China. Psychiatry Res 2016;241:55–60.

[10] Kroenke K, Evans E, Weitlauf S, McCalley S, Porter B, Williams T, et al. Comprehensive vs. Assisted Management of Mood and Pain Symptoms (CAMMPS) trial: study design and sample characteristics. Contemp Clin Trials 2018;64:179–87.

[11] Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1–3):163–73.

[12] Kroenke K, Spitzer RL, Williams JB, Lowe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345–59.

[13] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–13.

[14] Lowe B, Decker O, Muller S, Brahler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008;46(3):266–74.

[15] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63(11):1179–94.

[16] Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment 2011;18(3):263–83.

[17] Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, et al. Pain and its treatment in outpatients with metastatic cancer. N Engl J Med 1994;330(9):592–6.

[18] Toussaint A, Kroenke K, Baye F, Lourens S. Comparing the Patient Health Questionnaire - 15 and the Somatic Symptom Scale - 8 as measures of somatic symptom burden. J Psychosom Res 2017;101:44–50.

[19] Ware Jr. J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220–33.

[20] Ware Jr. JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473–83.

[21] Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996;11(Suppl. 3):89–95.

[22] Perkins AJ, Kroenke K, Unutzer J, Katon W, Williams Jr. JW, Hope C, et al. Common comorbidity scales were similar in their ability to predict health care costs and mortality. J Clin Epidemiol 2004;57(10):1040–8.

[23] Narasimhan M, Kannaday MH. Treating depression and achieving remission. Asian J Psychiatr 2010;3:163–8.

[24] de Zwart PL, Jeronimus BF, de Jonge P. Empirical evidence for definitions of episode, remission, recovery, relapse and recurrence in depression: a systematic review. Epidemiol Psychiatr Sci 2018:1–19.

[25] Montoya A, Lebrec J, Keane KM, Fregenal I, Ciudad A, Moríñigo Á, et al. Broader conceptualization of remission assessed by the remission from depression questionnaire and its association with symptomatic remission: a prospective, multi-center, observational study. BMC Psychiatry 2016;16:352.

[26] Johns SA, Kroenke K, Krebs EE, Theobald DE, Wu J, Tu W. Longitudinal comparison of three depression measures in adult cancer patients. J Pain Symptom Manage 2013;45(1):71–82.

[27] Kean J, Monahan PO, Kroenke K, Wu J, Yu Z, Stump TE, et al. Comparative responsiveness of the PROMIS pain interference short forms, brief pain inventory, PEG, and SF-36 bodily pain subscale. Med Care 2016;54(4):414–21.

[28] Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Med Care 2010;48(11):1007–14.

[29] Kroenke K, Theobald D, Wu J, Tu W, Krebs EE. Comparative responsiveness of pain measures in cancer patients. J Pain 2012;13(8):764–72.

[30] Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000;55(1):5–14.

[31] Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013;13:119.

[32] VanderWeele TJ. On the promotion of human flourishing. Proc Natl Acad Sci U S A 2017;114(31):8148–56.