Published in final edited form as:

Gen Hosp Psychiatry. 2016 ; 42: 60–66. doi:10.1016/j.genhosppsych.2016.07.005.

Performance of depression rating scales in patients with chronic kidney disease: an item response theory-based analysis^{★,★★,★,★★}

Marisa Toups, M.D.^a, Thomas Carmody, Ph.D.^{a,b}, Madhukar H. Trivedi, M.D.^a, A. John Rush, M.D.^c, and S. Susan Hedayati, M.D., M.H.Sc.^{d,e,*}

^aDepartment of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, USA, 75390-9119

^bDivision of Biostatistics, Department of Clinical Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, USA, 75390-9066

^cDivision of Clinical Sciences, Duke-NUS, 8 College Rd, Singapore, 169857

^dDivision of Nephrology, Department of Medicine, Veterans Affairs North Texas Health Care System, 4500 S. Lancaster Rd MC 111G1, Dallas, TX, USA, 75216

^eDivision of Nephrology, Department of Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, USA, 75390-8516

Abstract

Objective—Because there is overlap between somatic symptoms of depression and symptoms of chronic kidney disease (CKD), it is unclear if self-reported depression rating scales can be used accurately in predialysis CKD patients, especially if CKD and other comorbidities are symptomatic. We assessed the performance of two depression scales — the Beck Depression Inventory (BDI) and the Quick Inventory of Depression Symptomatology (QIDS-SR₁₆) — by CKD stage, diagnosis of diabetes and total medical comorbidity burden —using item response theory (IRT) in a sample of 272 predialysis CKD patients.

Methods—We performed IRT by low versus high CKD stage, diabetes versus no diabetes and high (>3 diagnoses) versus low medical comorbidity burden.

[★]Funding: This work was supported by a Veterans Affairs MERIT grant (CX000217–01) and a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, NIH) (R01DK085512) (Susan Hedayati, M.D.), the NIMH (K23MH104768) (Marisa Toups, MD), the Hershey Foundation and the Center for Depression Research and Clinical Care (Madhukar H. Trivedi, M.D). The views expressed here are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the NIH.

^{★★}Declaration of interest statement: The authors have no competing interests.

[★]Financial Disclosures: A. John Rush, M.D. has received consulting fees from Brain Resource Ltd., H. Eli Lilly, Lundbeck A/S, Medavante, Inc.; National Institute of Drug Abuse, Santium Inc., Takeda USA.; speaking fees from the University of California at San Diego, Hershey Penn State Medical Center, and the American Society for Clinical Psychopharmacology; royalties from Guilford Publications and the University of Texas Southwestern Medical Center; a travel grant from CINP and research support from Duke-National University of Singapore. Madhukar H. Trivedi, MD is or has been an advisor/consultant and received fees from: Alkermes, AstraZeneca, Cerecor, Eli Lilly & Company, Lundbeck, Naurex, Neuronetics, Otsuka Pharmaceuticals, Pamlab, Pfizer Inc., SHIRE Development and Takeda. Marisa Toups, MD has received fees from Otsuka Pharmaceuticals.

^{★★}Clinical Trial Registry: N/A

*Corresponding author at: VA North Texas Health Care System, Nephrology Section, 4500 South Lancaster Road MC 111G1, Dallas, TX, 75216-7167. Tel.: +1-214-857-2214; fax: +1-214-857-1514. susan.hedayati@utsouthwestern.edu (S.S. Hedayati).

Results—IRT models of each rating scale were affected in a limited way by CKD stage, diabetes and medical comorbidity burden. Sleep disturbances on the QIDS-SR₁₆ were more discriminatory for depression in diabetics and those with high comorbidity burden. Pessimism and guilt from the BDI compared to QIDS-SR₁₆ were more discriminatory of depression in the high CKD and high comorbidity groups, respectively.

Conclusions—Overall item differences were modest, and chronic disease severity by CKD stage, diabetes mellitus or other medical comorbidities did not appreciably contribute to differences in scale performance.

Keywords

Depression; Chronic kidney disease; Comorbidity; Screening; Assessment

1. Introduction

Prevalence of depression is increased in the setting of chronic kidney disease (CKD), such that 20–25% of patients are affected [1–3]. Importantly, depression is associated not only with significant functional impairment but also increased risk of hospitalization, disease progression and death in CKD patients [4,5]. A recent meta-analysis of 22 studies estimated depression increased risk of all-cause mortality in CKD patients by 60% [6]. Most studies analyzing prevalence or sequelae of depression in CKD patients used self-report questionnaires for identification and/or measurement [7,8]. Since somatic symptoms of depression such as decreased concentration, energy, appetite and sleep disturbance overlap with symptoms of CKD, self-reports may not perform as expected and lead to an overestimation of depression diagnoses in CKD patients. In fact, studies suggest that use of self-reports in CKD patients receiving dialysis or at dialysis initiation requires higher cutoff scores for accurate measurement of depression [9–11]. Our previous work demonstrated that in nondialysis CKD patients, cutoff scores for the diagnosis of a major depressive episode (MDE) were equivalent to previously established benchmarks and did not require adjustment [12]. In that analysis, we did not examine clinical differences within the nondialysis population, however, so it remains unknown whether illness severity —of CKD itself and due to comorbid illness — may affect the performance of depression self-report measures.

Item response theory (IRT) is a robust method to assess the performance of rating instruments in different samples [13]. IRT models the response characteristics of each item as well as the scale as a whole and, thus, provides a detailed evaluation of performance [14]. To our knowledge, neither the Quick Inventory of Depression Symptomatology Self-Rated (QIDS-SR₁₆) nor the Beck Depression Inventory (BDI) have previously been analyzed using IRT in nondialysis CKD patients, though the BDI has been the subject of IRT analysis in other medical populations. Comparing patients who were above age 65 to younger adults IRT analysis showed that older patients reported more somatic symptoms on the BDI, a trend that was more pronounced as depression severity increased [15]. However, the authors did not attempt to account for the role of physical health in their results. Wardenaar et al. performed a similar IRT analysis of the BDI in subjects postmyocardial infarction and found that mood items were more representative of diagnosed depression, compared to somatic items, which were reported frequently in all subjects [16]. Collectively, these analyses

support the use of IRT to better understand the effect of medical comorbidity on the assessment of depression.

Here, we report upon an IRT analysis of the BDI and QIDS-SR₁₆ in a large, consecutively recruited cohort with nondialysis Stage 2–5 CKD. The specific objectives were to determine whether CKD stage and presence of other medical comorbidities impact scale performance and, if so, which item(s) account for the differences. Based on previous findings, we would expect that somatic symptoms would drive any difference found in scale performance as disease severity increases.

2. Methods

This study was approved by the Institutional Review Boards of UT Southwestern Medical Center and the North Texas Veterans Affairs Medical Center.

2.1. Participants

This sample was recruited for a prospective observational study designed to identify the prevalence and effects of major depressive disorder (MDD) in the predialysis CKD population [5] and has been reported on previously [12,17]. Participants were adult outpatients with CKD Stages 2, 3, 4 and Predialysis 5, attending the Dallas Veterans Affairs Medical Center Nephrology Clinic, recruited in 2005–2006. Patients with Stage 1 CKD were excluded. Participants were approached consecutively in the clinic waiting area using a schema in which, to avoid selection bias, every sixth patient deemed potentially eligible based on medical record review was chosen for recruitment. All participants completed written, informed consent prior to providing any information to investigators. All subjects receive a formal psychiatric diagnostic interview using the MINI International Neuropsychiatric Interview (MINI) in addition to providing self-assessment of depression symptoms [18].

Of the 272 participants, 5 did not have complete diagnostic data and were excluded, leaving 267. A partially overlapping group was missing items on either the QIDS-SR₁₆, BDI or both and was excluded from the respective analyses. Therefore, 263 participants were eligible for the QIDS-SR₁₆ analysis and 252 for the BDI analysis.

2.1.1. Group definitions—We divided the participants into low (2 or 3) and high stage (4 or 5) CKD. Stage 2–3 CKD is mild-to-moderate and asymptomatic, while Stages 4–5 is considered severe and symptomatic [19]. CKD stage was defined at baseline using the National Kidney Foundation guidelines [19]: Stage 2, estimated glomerular filtration rate (eGFR) 60–89 mL/min/1.73 m² and other evidence of kidney disease manifest by either pathologic abnormality of kidney on biopsy or markers of kidney damage; Stage 3, eGFR 30–59; Stage 4, eGFR 15–29; and Stage 5, eGFR<15. From the 267 participants who had data on CKD stage and at least one depression scale, 118 (44.2%) had low stage and 149 (55.8%) had high stage CKD.

We chose diabetes for individual analysis because it is the most common cause of CKD in the United States [20] and was the most common comorbidity (55.6%) in our sample.

Diabetes was defined by medical record diagnosis of either Type I or II diabetes or use of insulin or other medications indicated for normalizing blood sugar. By this definition, diabetes was present in 148 participants (55.6%) from the 266 who had data on diabetes diagnosis and at least one depression scale.

We used an a priori definition of *high burden* of medical comorbidity as >3 and *low burden* as ≤ 3 diagnoses present in the medical record for each individual subject. The threshold was chosen pragmatically; there is no consensus on the definition of medical burden in the context of depression diagnosis and treatment [21]. The median number of comorbidities was 3; therefore, a split at 3 produced two groups of comparable size. Tracked medical comorbidities included hypertension, diabetes mellitus, congestive heart failure, coronary artery disease, cerebrovascular and peripheral vascular disease, lung disease, liver disease, nonskin malignancy and infection with human immunodeficiency virus. Each diagnostic category was used only once; for example, if a patient had two types of malignancy, this was counted as one comorbidity. Of the 267 participants with data from at least one scale and complete medical record data, 111 (41.6%) had more than three medical comorbidities and were classed as high burden.

2.2. Depression assessments

The BDI [22] has 21 items rated using a 0–3-point Likert scale and summed for the final score, with a maximum of 63. Higher scores correspond to more severe symptoms. The QIDS-SR₁₆[23] contains sixteen items covering the nine DSM-IV symptom domains for MDD. Three domains have multiple items: sleep has four items rating initial, middle and terminal insomnia and hypersomnia; appetite has four items rating increased appetite, decreased appetite and increased and decreased weight; and psychomotor has items assessing increased and decreased psychomotor activity. Only the highest rated item in each of these domains is used in the final score, for a range of 0–27. For this analysis, only the scored item was used in the IRT model; therefore, the magnitude of the dysfunction in those domains is modeled rather than the specific symptom profile.

2.3. Statistical analysis

Baseline demographic and clinical characteristics were compared between groups using *t* tests for continuous outcomes and chi-square tests for categorical outcomes.

Before applying IRT to a sample of assessment scores, it first must be shown that the scale(s) of interest are unidimensional in the analyzed sample. To do this, we used exploratory factor analysis based on polychoric correlations and parallel analysis (PA), as modified by Glorfeld, to determine the number of factors [24]. Simulation studies support the accuracy of this method over other methods, such as the scree test, to determine the best number of factors [24,25]. PA is conducted by generating simulated random data sets with the same number of items and observations as the real data set. The eigenvalues derived from these datasets determine the size of eigenvalues which may occur by chance. We created 5000 simulated data sets and took the 95th percentile of the eigenvalue distribution as the threshold for the eigenvalues derived from the real data. If only one factor derived

from the real data set has an eigenvalue larger than this threshold, the scale is considered to be unidimensional.

Next, IRT was applied to each scale and for each condition using MULTILOG Version 7 software [26] to run a four-parameter model following Samejima et al. [27]. Items from both scales have four ordered response categories which represent levels of an underlying continuous distribution. As the model of Samejima et al. is based on the assumption of an underlying continuous distribution, we consider this model more appropriate than other models for ordered response categories such as the partial credit model [28]. The parameters used were **a** - the slope, where a higher value indicates more discrimination between lower and higher levels of depression, and three threshold parameters, **b₀**, **b₁** and **b₂**. Each threshold parameter defines the depression severity at which there is equal probability that a participant will choose between a lower versus a higher scored option on that item. Thus, **b₀** describes the threshold between scores of 0 and 1–3, **b₁** describes the threshold between scores of 0–1 and 2–3 and **b₂** describes the threshold dividing 0–2 from 3. As the **b** parameter values increase, a higher level of depression is required for 50% chance rating the item above each threshold. Depression severity is expressed in a normalized scale with mean 0 and standard deviation 1. So, for example, if **b₁** for a particular item is 2.5, a participant with a level of depression 2.5 standard deviations above the mean had equal chance of endorsing a score of 0–1 versus 2–3 on that item. If two groups have significantly different parameters for an item, factors other than depression are assumed to be driving the difference, called *differential item functioning* (DIF).

We tested for DIF using a goodness of fit analysis of G^2 (the negative of the twice the log likelihood statistic). In each case, a constrained version of the model in which some or all of the parameters are forced into equality between experimental groups (e.g., diabetes vs. no diabetes) was compared to a less constrained model using the difference in G^2 which is distributed as a chi-square statistic. The constrained condition models the scenario in which the two groups respond identically; a fully unconstrained model represents the best overall fit of the model to the data. If these two models show difference in fit, then the hypothesis that the groups rate items identically can be rejected. In addition to constraining all parameters, we performed models for which all slope and all location parameters were respectively constrained, as well as constraining parameters by item, to evaluate which item(s) contributed to any observed difference in fit. To control for the large number of tests, the procedure of Hommel (1989) was used to adjust the significance threshold for the individual item tests, using a threshold of $\alpha=0.05$ for the nine items of the QIDS-SR₁₆ and the 21 items of the BDI, respectively. Multiple imputation was not considered necessary as only 1.5% and 5.6% of participants had missing items on the QIDS-SR₁₆ and the BDI, respectively.

3. Results

3.1. Baseline characteristics

Descriptive statistics of the sample as a whole and by CKD stage groups are given in Table 1. When comparing participants with low stage to those with high stage CKD, there were significant differences in CKD stage-associated variables, but demographic variables and

scores on the BDI and QIDS-SR₁₆ were not significantly different. When comparing participants with and without diabetes, more differences were found including higher QIDS-SR₁₆ (but not BDI) score ($P=.013$) and higher prevalence of MDE ($P=.018$) (Supplemental Table 1). Comparing participants with high versus low burden, we found that BDI and QIDS-SR₁₆ scores were higher in the high burden group ($P<.008$ and 0.002 , respectively). A larger proportion of the high burden group had a MINI diagnosis of depression ($P=.009$) and diabetes mellitus ($P<.001$) (Supplemental Table 2). Fig. 1 shows the mean score on each of the nine QIDS-SR₁₆ domains, plotted by group, providing a visual reference for relative mean scores in each domain. Data for the BDI items scores were similar (not shown).

3.2. Factor analysis

PA showed the QIDS-SR₁₆ to be unidimensional in our sample. The largest real data eigenvalue (5.0) exceeded the 95th percentile from the simulated distribution (1.4), while the second data eigenvalue (0.9) was smaller than the 95th percentile of the simulated second eigenvalues (1.2). For the BDI, the 95th percentile for the largest simulated eigenvalue was 1.7 and the largest real data eigenvalue was 11.4, but the second data eigenvalue (1.6) slightly exceeded the 95th percentile of the simulated distribution (1.5) which suggests two factors. However, because the first eigenvalue was much larger than the second, and these two factors were highly correlated ($r=0.78$), we determined that one factor was dominant, and therefore, the BDI is sufficiently unidimensional in our sample to proceed with the IRT analysis [14].

3.3. IRT results

Supplemental Tables 3 and 4 show IRT parameter for the QIDS-SR₁₆ and the BDI for the sample as a whole. Table 3 shows that there was a range of 1.07 to 2.48 for the **a** or slope parameters on the QIDS-SR₁₆ with sleep being the least discriminative item and sad mood the most. Sleep problems were endorsed with high probability, as depression severity -2.98 standard deviations below the sample mean was associated with 50% chance of endorsing any sleep problem. The item assessing suicidal ideation was the least likely to be endorsed, with a **b**₀ threshold of 1.80. On the BDI, suicidality was the most discriminative item with a slope of 4.70 and also the least likely to be endorsed for a given level of depression **b**₀=1.51 (Table 2). Fatigue was the symptom most likely to be endorsed with a **b**₀= -1.12 , followed by loss of energy, **b**₀= -0.41 .

3.3.1. Comparison by CKD stage—Comparison of the fully constrained and fully unconstrained models of QIDS-SR₁₆ across CKD stage groups resulted in a significant difference in fit ($G^2=54.9$, $df=36$, $P=.023$). Constraining only the slope parameters or only the threshold parameters did not result in significant difference in fit. The analysis of each item's parameters separately also showed no significant differences after adjustment for multiple comparisons. Although significant overall DIF was found, it was not possible to characterize because it did not localize to slope or threshold parameters or to specific items (Supplemental Table 5).

Similar analysis for the BDI by CKD stage showed no significant differences in fit for the comparison between a nonconstrained and a fully constrained model. When constraining

only the slope parameters, there was a significant difference in fit ($G^2=35.2$, $df=21$, $P=.027$), but not when constraining only threshold parameters. In the individual item analyses, Item 2 (pessimism) showed a significant decrease in fit when constraining only the slope parameter ($G^2=10.2$, $df=1$, $P=.029$). This indicates that pessimism accounts for the overall difference in fitting the BDI slope parameters and that higher scores on this item are more discriminatory (more indicative of level of depression) in the high stage group ($a_{\text{high}}=4.82$, $a_{\text{low}}=1.65$) (Table 2).

3.3.2. Comparison by presence of diabetes mellitus—The analysis of the QIDS-SR₁₆ scores for CKD patients divided according to the presence of diabetes (Supplemental Table 6) showed significant difference in fit between the models with unconstrained and fully constrained parameters ($G^2=58.6$, $df=36$, $P=.01$). Constraining only the slope parameters did not significantly change the fit from the unconstrained model but constraining threshold parameters did ($G^2=43.9$, $df=27$, $P=.021$); therefore, the difference in fit is accounted for by the threshold parameters. When modeling each item individually, the sleep item (Item 1) showed significantly different fit after adjusting for multiple comparisons when both slope and threshold parameters were constrained ($G^2=15.2$, $df=4$, $P=.039$). In patients with diabetes score on, the sleep item is more correlated to level of depression than for non-diabetic patients ($a_{\text{DM}}=1.15$, $a_{\text{noDM}}=0.99$), and patients with diabetes ($b_0=-4.21$, $b_1=-1.97$, $b_2=0.26$) were more likely to endorse sleep problems than patients without diabetes ($b_0=-2.36$, $b_1=-1.53$, $b_2=0.81$) given the same level of depression.

In the analysis of BDI responses by presence or absence of diabetes, there was no significant difference found in any comparison (Supplemental Table 7).

3.3.3. Comparison by comorbidity burden—Comparing the fully unconstrained model of QIDS-SR₁₆ responses with a fully constrained model did not result in significant difference in fit. Neither did comparing models in which only the slope or only the threshold parameters were constrained with the fully unconstrained model. When constraining parameters by item, QIDS-SR₁₆ Item 1 (sleep) showed significant difference in fit constraining both its slope and threshold parameters ($G^2=20.0$, $df=4$, $P=.004$) as well as the threshold parameters only ($G^2=19.8$, $df=3$, $P=.002$). A higher slope parameter in the high burden group suggests that sleep item score is more discriminatory for those patients ($a_{\text{High}}=1.33$, $a_{\text{Low}}=0.87$). The threshold parameters for the high burden group were $b_0=-2.98$, $b_1=-1.45$ and $b_2=-0.08$, while those for the low burden group were $b_0=-3.19$, $b_1=-2.17$ and $b_2=-1.04$ (Supplemental Table 8). For a given depression severity, the low burden group was more likely to complain of mild sleep problems; patients with high burden were more likely to endorse a score of 3.

Modeling responses on the BDI showed no significant difference in fit for the scale-wide analyses (Supplemental Table 9). When modeling by item, no items showed significant difference in fit when both slope and threshold parameters were constrained or when slope parameters only were constrained. However, when constraining threshold parameters by item, Item 5 (guilt) showed significant difference in fit ($G^2=22.9$, $df=3$, $P<.001$). Higher threshold parameter values for the low burden group indicated that more severe depression was required for endorsement of guilt in these patients as compared to the high burden group.

($b_0=0.78$, $b_1=1.21$ and $b_2=2.09$ for high burden and $b_0=1.64$, $b_1=2.75$ and $b_2=7.49$ for low burden).

4. Discussion

To our knowledge, this is the first IRT analysis of the QIDS-SR₁₆ in outpatients with predialysis CKD and the first of the BDI by stage and medical comorbidities, including specifically diabetes. We set out to identify whether disease severity within the population of CKD patients affects the performance of common depression self-report scales. We found that the BDI items performed according to a pattern previously identified in medically ill samples, in which items assessing mood and negative thinking symptoms were less likely to be endorsed without significant depression and more discriminatory of depression than items assessing psychomotor symptoms, which were likely to be endorsed regardless of depression severity. The QIDS-SR₁₆ items on the other hand were more likely to be endorsed at low or moderate levels of depression. In the subgroup comparisons, the QIDS-SR₁₆ and BDI showed no widespread differences in performance based on CKD stage, presence of diabetes or comorbidity burden. We did not see the relative over endorsement of somatic symptoms increase in the participants with greater medical illness severity. Therefore, clinicians may interpret the scale results within the CKD population without concern that scores are greatly affected by the burdens of worsening CKD or related medical illnesses.

Our findings for the QIDS-SR₁₆ demonstrated that sleep complaints had the highest probability of endorsement of the nine domains in the sample as a whole. Sizable negative threshold parameters for the QIDS-SR₁₆ sleep item show that even patients with depression severity multiple standard deviations below the mean for the sample were still likely to endorse some level of insomnia or hypersomnia. This is not surprising as poor sleep quality and primary sleep disorders are both common in CKD patients [29,30]. It may therefore seem paradoxical that sleep was more able to discriminate depression for patients with diabetes versus no diabetes and high burden versus low burden. This may be interpreted as an assurance that the unusual CKD patient free of sleep disturbance is not depressed. Or it may mean that severe problems — 3 on a QIDS-SR₁₆ item — are necessary for concern for depression, while low scores for sleep problems should be attributed to medical causes. In particular for the burden analysis, the pattern of threshold parameters shows that patients with less burden were more likely to endorse mild sleep problems but less likely to endorse the maximum score than participants with high burden. This suggests that it is the severe sleep complaints that drive the difference between burden groups. Lastly, we note that, as compared to the BDI, which asks only about insomnia, the QIDS-SR₁₆ includes hypersomnia in this domain. This may account for the lack of replication of sleep item performance differences on the BDI.

Finally, symptoms of negative thinking on the BDI may be more indicative of depression, especially as patients become more ill. High stage CKD was associated with more reported pessimism and high burden was associated with more reported guilt. Our finding is supported by a previous factor analysis of the BDI in dialysis-dependent CKD; Chilcot et al. found that a core depression factor (which included the guilt and pessimism items) accounted for 81% of the variance in symptoms. Meanwhile, orthogonal cognitive and

somatic factors only explained 8% and 9% of the common variance, respectively [31]. The focus of the BDI on negative thinking may support the use of BDI in CKD and other medical illnesses if these symptoms are best suited to tracking depression across the spectrum of illness severity. In fact, Beck designed a version of the BDI for use in primary care which includes the pessimism item [32]. The relative exclusion of this type of symptom from the QIDS-SR₁₆ may account for the failure to replicate this BDI finding.

4.1. Limitations

Our factor analysis showed that the BDI was not entirely unidimensional in this sample, which could have affected the IRT results. In addition, in a few cases, significant difference in overall fit was found — such as for the QIDS-SR₁₆ by CKD stage — but without association with any individual items. Because IRT analysis is not adjustable, factors which differed across the groups, such as age, may account for these significant but uncategorizable differences. We found that diabetes and comorbidity burden were associated with higher rates of MINI diagnosed depression, making the distribution of depression across our comparison groups uneven and perhaps biasing the results. Finally, the sample was comprised predominantly of men, and our results may not be generalizable to women with CKD and depression.

4.2. Conclusions

The BDI and QIDS-SR₁₆ depression self-reports effectively measure depressive symptoms in patients with Stages 2–5 nondialysis CKD without large effects due to illness severity of CKD or common comorbidities. Neither CKD stage, presence of diabetes mellitus nor number of medical comorbidities caused widespread changes in item performance. Sleep reports on the QIDS-SR₁₆ and negative thinking on the BDI may reflect progression of medical illness burden toward the pattern recognized in dialysis patient in which somatic symptoms are less sensitive for the assessment of depression. However, until dialysis initiation is imminent, increasing illness burden should not interfere with interpretation of self-reported depression symptoms by clinicians in practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the assistance of Jessica Harper in preparing the tables and manuscript.

References

1. Palmer S, Vecchio M, Craig JC, Tonelli M, Johnson DW, Nicolucci A, et al. Prevalence of depression in chronic kidney disease: systematic review and meta-analysis of observational studies. *Kidney Int.* 2013; 84:179–91. [PubMed: 23486521]
2. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care.* 2001; 24:1069–78. [PubMed: 11375373]
3. Rutledge T, Reis VA, Linke SE, Greenberg BH, Mills PJ. Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *J Am Coll Cardiol.* 2006; 48:1527–37. [PubMed: 17045884]

4. Hedayati SS, Bosworth HB, Briley LP, Sloane RJ, Pieper CF, Kimmel PL, et al. Death or hospitalization of patients on chronic hemodialysis is associated with a physician-based diagnosis of depression. *Kidney Int.* 2008; 74:930–6. [PubMed: 18580856]
5. Hedayati SS, Minhajuddin AT, Afshar M, Toto RD, Trivedi MH, Rush AJ. Association between major depressive episodes in patients with chronic kidney disease and initiation of dialysis, hospitalization, or death. *JAMA.* 2010; 303:1946–53. [PubMed: 20483971]
6. Palmer SC, Vecchio M, Craig JC, Tonelli M, Johnson DW, Nicolucci A, et al. Association between depression and death in people with CKD: a meta-analysis of cohort studies. *Am J Kidney Dis.* 2013; 62:493–505. [PubMed: 23623139]
7. Cohen SD, Patel SS, Khetpal P, Peterson RA, Kimmel PL. Pain, sleep disturbance, and quality of life in patients with chronic kidney disease. *Clin J Am Soc Nephrol.* 2007; 2:919–25. [PubMed: 17702733]
8. Shidler NR, Peterson RA, Kimmel PL. Quality of life and psychosocial relationships in patients with chronic renal insufficiency. *Am J Kidney Dis.* 1998; 32:557–66. [PubMed: 9774115]
9. Hedayati SS, Bosworth HB, Kuchibhatla M, Kimmel PL, Szczech LA. The predictive value of self-report scales compared with physician diagnosis of depression in hemodialysis patients. *Kidney Int.* 2006; 69:1662–8. [PubMed: 16598203]
10. Watnick S, Wang PL, Demadura T, Ganzini L. Validation of 2 depression screening tools in dialysis patients. *Am J Kidney Dis.* 2005; 46:919–24. [PubMed: 16253733]
11. Craven JL, Rodin GM, Littlefield C. The Beck depression inventory as a screening device for major depression in renal dialysis patients. *Int J Psychiatry Med.* 1988; 18:365–74. [PubMed: 3235282]
12. Hedayati SS, Minhajuddin AT, Toto RD, Morris DW, Rush AJ. Validation of depression screening scales in patients with CKD. *Am J Kidney Dis.* 2009; 54:433–9. [PubMed: 19493600]
13. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther.* 2014; 36:648–62. [PubMed: 24811753]
14. Hambleton, RKSH., Rogers, HJ. *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications; 1991.
15. Kim Y, Pilkonis PA, Frank E, Thase ME, Reynolds CF. Differential functioning of the Beck depression inventory in late-life patients: use of item response theory. *Psychol Aging.* 2002; 17:379–91. [PubMed: 12243380]
16. Wardenaar KJ, Wanders RB, Roest AM, Meijer RR, De Jonge P. What does the beck depression inventory measure in myocardial infarction patients? A psychometric approach using item response theory and person-fit. *Int J Methods Psychiatr Res.* 2015; 24:130–42. [PubMed: 25994207]
17. Hedayati SS, Minhajuddin AT, Toto RD, Morris DW, Rush AJ. Prevalence of major depressive episode in CKD. *Am J Kidney Dis.* 2009; 54:424–32. [PubMed: 19493599]
18. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998; 59(Suppl 20): 22–33. [quiz 34–57].
19. National KF. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis.* 2002; 39:S1–266. [PubMed: 11904577]
20. United States Renal Data System, Services DoHaH. *USRDS annual data report.* Bethesda, MD: 2007.
21. Menear M, Dore I, Cloutier AM, Perrier L, Roberge P, Duhoux A, et al. Chronic physical comorbidity burden and the quality of depression treatment in primary care: a systematic review. *J Psychosom Res.* 2015; 78:314–23. [PubMed: 25649274]
22. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961; 4:561–71. [PubMed: 13688369]
23. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report

- (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003; 54:573–83. [PubMed: 12946886]
24. Glorfeld L. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educ Psychol Meas*. 1995; 55:377–93.
25. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychol Bull*. 1986; 99:432–42.
26. Thissen, D., Chen, WH., Block, RD. MULTILOG 7 for windows: multiple-category item analysis and test scoring using item response theory. Skokie, IL: Scientific Software International, Inc; 2003.
27. Samejima, F., Linden, W. Graded response model. In: Linden, HR., editor. *Handbook of modern item response theory*. New York, NY: Springer-Verlag; 1997. p. 85-100.
28. Baker JG, Rounds JB, Zevon MA. A comparison of graded response and rasch partial credit models with subjective well-being. *J Educ Behav Stat*. 2000; 25:253–70.
29. Plantinga L, Lee K, Inker LA, Saran R, Yee J, Gillespie B, et al. Association of sleep-related problems with CKD in the United States, 2005–2008. *Am J Kidney Dis*. 2011; 58:554–64. [PubMed: 21816524]
30. Unruh M, Tamura MK. Introduction: sleep and neurologic disorders in chronic kidney disease. *Semin Nephrol*. 2015; 35:303. [PubMed: 26355248]
31. Chilcot J, Norton S, Wellsted D, Almond M, Davenport A, Farrington K. A confirmatory factor analysis of the Beck depression inventory-II in end-stage renal disease patients. *J Psychosom Res*. 2011; 71:148–53. [PubMed: 21843749]
32. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck depression inventory for primary care. *Behav Res Ther*. 1997; 35:785–91. [PubMed: 9256522]

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.genhosppsych.2016.07.005>.

Fig. 1.
A–C. Mean reported score by item for the QIDS-SR₁₆ for each set of analysis groups. (A) CKD stage; (B) diabetes diagnosis; and (C) comorbidity burden.

Table 1

Baseline parameters for all participants and low stage (2 or 3) versus high stage (4 or 5) CKD.

Baseline variable	Analyzed sample mean (S.D.) or <i>N</i> (%) (<i>N</i> =267)	Low stage CKD mean (S.D.) or <i>N</i> (%) (<i>N</i> =118)	High stage CKD mean (S.D.) or <i>N</i> (%) (<i>N</i> =149)	Low stage vs. high stage <i>P</i> - value
Age (years)	64.42 (12.0)	63.96 (12.1)	64.79 (11.9)	0.576
No. comorbid illnesses	3.24 (3.24)	3.16 (1.7)	3.31 (1.4)	0.435
Hemoglobin (g/dL)	12.45 (2.0)	13.00 (2.0)	12.01 (1.9)	<0.001
Ca × Phos Product (mg ² /dL ²) (mg ² /dL ²) (mg ² /dL ²)v(mg ² /dL ²) (mg ² /dL ²) (mg ² /dL ²)	36.62 (8.8)	32.76 (5.9)	39.68 (9.4)	<0.001
Albumin (g/dL)	4.00 (0.5)	4.03 (0.5)	3.98 (0.5)	0.334
Creatinine (mg/dL)	3.22 (2.4)	2.13 (2.3)	4.08 (2.2)	<0.001
Blood urea nitrogen asddas(mg/dL)	42.19 (20.5)	29.70 (10.8)	52.07 (21.0)	<0.001
eGFR (mg/ml/1.73m ²) (mL/min/1.73m ²) (mL/min/1.73 m ²)	31.55 (16.7)	45.95 (14.2)	20.14 (6.7)	<0.001
BDI total score	7.82 (7.8)	7.56 (6.8)	8.01 (8.4)	0.649
QIDS-SR total score	7.07 (4.9)	6.64 (4.9)	7.40 (4.9)	0.218
Male gender	265 (99.3)	117 (99.2)	148 (99.3)	0.868
Race				0.384
White	151 (56.6)	63 (53.4)	88 (59.1)	
African-American	99 (37.1)	45 (38.1)	54 (36.2)	
Other	17 (6.4)	10 (8.5)	7 (4.7)	
Completed high school	155 (59.4)	76 (65.5)	79 (54.5)	0.071
Married	179 (67.3)	77 (65.8)	102 (68.5)	0.648
Employed	51 (19.3)	27 (23.1)	24 (16.2)	0.159
Diabetes mellitus	148 (55.6)	61 (51.7)	87 (58.8)	0.248
Hypertension	259 (97.0)	111 (94.1)	148 (99.3)	0.012
Cardiovascular disease	168 (63.2)	72 (61.5)	96 (64.4)	0.628
History drug/alcohol abuse	79 (29.8)	35 (29.9)	44 (29.7)	0.974
History of depression	65 (24.8)	33 (28.5)	32 (21.9)	0.224
Taking antidepressants	53 (19.9)	26 (22.0)	27 (18.1)	0.426
Current MDE	56 (20.97)	23 (19.5)	33 (22.2)	0.597

Table 2

BDI IRT parameters for high stage (4 or 5) versus low stage (2 or 3) CKD.

BDI item	Stage CKD	a	b ₀	b ₁	b ₂
Sadness	High	2.99	0.62	1.29	2.41
	Low	2.97	0.73	1.55	2.51
Pessimism [†]	High	4.82	0.98	1.32	1.99
	Low	1.65	1.06	2.29	3.50
Failure	High	4.08	0.98	1.68	13.00
	Low	2.69	1.20	1.74	5.35
Dissatisfaction	High	3.63	0.37	1.99	2.29
	Low	1.91	0.55	2.49	16.35
Guilt	High	2.91	1.04	1.49	2.41
	Low	3.38	1.17	1.90	4.46
Punishment	High	2.56	1.37	2.08	6.87
	Low	2.82	1.49	2.27	5.49
Self-dislike	High	5.24	0.80	1.96	*
	Low	2.05	0.99	13.07	*
Self-critical	High	3.71	1.14	1.64	4.19
	Low	2.94	1.00	2.54	4.90
Suicide	High	4.71	1.43	1.70	3.56
	Low	6.82	1.48	1.86	3.29
Crying	High	2.01	1.14	2.08	2.62
	Low	1.19	1.77	3.82	4.41
Irritation	High	2.14	0.57	2.51	*
	Low	1.44	0.53	3.34	*
Loss interest	High	2.22	1.01	1.89	3.00
	Low	3.07	0.67	1.92	7.88
Indecisiveness	High	2.61	1.00	2.13	11.79
	Low	1.86	1.55	2.56	3.17
Appearance	High	2.26	0.97	2.19	2.97
	Low	1.43	1.55	2.61	8.83

BDI item	Stage CKD	a	b ₀	b ₁	b ₂
Loss energy	High	1.91	-0.45	1.35	2.52
	Low	2.14	-0.36	1.11	2.20
Sleep	High	1.10	0.42	1.56	2.15
	Low	1.81	0.28	1.25	1.98
Fatigue	High	1.63	-1.27	1.12	2.38
	Low	1.67	-0.92	1.23	2.04
Appetite	High	1.45	0.85	2.31	3.14
	Low	1.46	1.10	2.79	3.79
Weight	High	1.07	1.47	2.27	2.59
	Low	0.67	1.91	4.52	5.17
Health	High	1.75	0.45	2.00	2.62
	Low	1.16	0.59	2.85	4.49
Sex	High	0.99	0.51	2.10	3.16
	Low	0.74	0.81	2.35	3.38

P = .029 slope only comparison.

P = .258 threshold only comparison.

* Insufficient data to compute an estimate.

[†] *P* = .099 slope and threshold comparison between high vs. low stage CKD.

Table 3

QIDS-SR¹⁹ IRT parameters for those with and without diabetes.

QIDS domain		a	b ₀	b ₁	b ₂
Sleep [†]	Diabetes	1.15	-4.21	-1.97	0.26
	No diabetes	0.99	-2.36	-1.53	0.81
Sad mood	Diabetes	2.79	0.21	1.00	1.62
	No diabetes	1.95	0.85	1.64	2.22
Appetite	Diabetes	0.94	-0.70	0.42	1.43
	No diabetes	1.59	-0.16	0.83	1.46
Concentration	Diabetes	2.34	0.47	1.24	2.51
	No diabetes	2.69	0.54	1.68	2.14
Self outlook	Diabetes	3.48	1.06	1.48	1.80
	No diabetes	1.47	1.56	2.57	2.88
Suicide	Diabetes	2.07	1.43	2.10	7.97
	No diabetes	0.89	3.15	4.14	14.27
Involvement	Diabetes	1.91	0.66	1.44	2.44
	No diabetes	1.29	0.84	2.17	26.59
Energy	Diabetes	1.76	-0.84	0.69	1.64
	No diabetes	1.49	-0.63	0.99	2.11
Psychomotor	Diabetes	1.78	0.23	1.29	2.24
	No diabetes	2.60	0.33	1.47	2.49

P = .584 slope only comparison.

P = .142 threshold only comparison.

[†] *P* = .039 slope and threshold comparison between presence vs. absence of diabetes.