

Casting a wider net in behavioral health screening in primary care: a preliminary study of the Outcome Rating Scale

Brian DeSantis¹, Miranda J. Jackson¹, Barry L. Duncan² and Robert J. Reese³

¹Peak Vista Community Health Centers, Colorado Springs, CO, USA

²Heart and Soul of Change Project, Jensen Beach, FL, USA

³University of Kentucky, Lexington, KY, USA

Introduction: The integration of behavioral health services into primary care has led to enhanced use of brief screening measures to identify mental health problems. Although useful, such instruments are largely symptom based and diagnosis specific. This narrow focus can potentially limit the identification of broader social or relational distress in patients that affect medical outcomes, as well as present feasibility challenges using a multi-measure approach in identifying mental health comorbidities. **Method:** This exploratory study of adult primary care patients compared an ultra-brief, and widely used measure of global distress across life functioning, the Outcome Rating Scale (ORS), with the Patient Health Questionnaire (PHQ-9 and PHQ-2). **Results:** Correlations between the ORS and the PHQ-9 and PHQ-2 indicated agreement between the measures in classifying patients, and the ORS identified significantly more patients in the clinical range. **Discussion:** Although results are preliminary, the ORS may cast a wider net in identifying patients with significant distress in primary care.

Key words: ORS; PHQ-9; primary care screens

Received 15 May 2016; revised 25 July 2016; accepted 7 August 2016

The majority of patients with mental health (MH) conditions are assessed and treated in primary care (Hunter *et al.*, 2009; Petterson *et al.*, 2014). With the emergence of integrated behavioral health (BH), brief screening measures are increasingly used to identify patients with MH problems and assist in interdisciplinary clinical decisions to improve patient care. Such screening tools typically focus on MH symptomatology associated with one disorder. Given that depression is the most prevalent MH condition with the majority of depressed adults receiving treatment in primary care (Bland, 2004; Edlund *et al.*, 2004), routine depression screening has been recommended by the US Preventive Services Task Force

(US Preventive Services Task Force, 2002; 2009; Siu and US Preventive Services Task Force, 2016). The US Department of Health and Human Services, Health Resources and Services Division has also recently required depression screening for patients seen in Federally Qualified Health Centers (FQHCs).

The first adult depression screen designed for primary care was the Patient Health Questionnaire-9 (PHQ-9) (Kroenke *et al.*, 2001), which was followed by a briefer PHQ-2 (Kroenke *et al.*, 2003). Both the PHQ-2 and PHQ-9 have demonstrated high sensitivity and specificity for detecting major depression in primary care (Arroll *et al.*, 2010) and a two-step approach to primary care depression screening is often utilized to improve diagnostic accuracy and assess severity. However, in a retrospective chart review at a university hospital-based family medicine clinic with integrated behavioral health providers (BHPs), only 5% of 200 family medicine patients

Correspondence to: Brian DeSantis, Behavioral Health, Peak Vista Community Health Centers, Pediatric Health Center, 2828 International Circle, Suite 140, Colorado Springs, CO 80910-3195, USA. Email: brian.desantis@peakvista.org

© Cambridge University Press 2016

with a positive PHQ-2 score were administered a PHQ-9, with physicians reporting competing demands, time constraints, and prior knowledge of their patient's depression status as reasons for not administering a follow-up PHQ-9 (Fuchs *et al.*, 2015). Thus, despite the validity of the PHQ-9 and its wide use in research trials, real world workflow demands challenge the extent to which it is being routinely used in primary care settings (Blasinsky *et al.*, 2006).

Besides the feasibility concerns of a stepped-method screening approach, symptom-based screening tools like the PHQ-9 are potentially limited by their diagnostic focus. They tend not to identify the broad array of MH distress (eg, relational, social) that bring patients to primary care that likely influence both emotional and physical health, nor can they detect common MH comorbidities. Thus, a singular use of many of these traditional screening measures might not identify a number of patients suffering from other MH symptoms or other life problems who could benefit from BH consultation. Although the use of multiple screening measures might solve the under-identification problem, comprehensive assessment is not practical given the workflow demands of primary care.

A brief global distress measure with a broad focus on life functioning may offer an alternative. The Outcome Rating Scale (ORS) (Miller and Duncan, 2000) is one of two measures comprising the Partners for Change Outcome Management System (PCOMS) (Duncan, 2012; Duncan and Reese, 2015). The ORS is an ultra-brief, validated visual analogue, self-report measure of a patient's perceived level of global distress and functioning across four life domains: individual, interpersonal, social, and overall. PCOMS was originally developed and researched as a feasible clinical and outcome system for specialty MH settings and is included in the Substance Abuse and Mental Health Administration's (SAMHSA) National Registry of Evidence-Based Programs and Practices (NREPP). Although the ORS measures and tracks patient change in MH and substance abuse services, it has never been investigated as a primary care BH screener.

The principle aim of this preliminary study was to investigate if a single measure of global distress in four life functioning domains could serve as a universal primary care screener. To serve this purpose, this exploratory study compared the ORS with the PHQ-9 and PHQ-2, evaluating their

correlations, reliability coefficients, and the number of patients who screened positive for potential BH consultation. We hypothesized that the ORS, which takes a more comprehensive picture of functioning beyond symptoms, would classify a higher percentage of patients who may benefit from consultation.

Method

Setting

This study was conducted at three small rural family practice health centers, associated with Peak Vista Community Health Centers, a large FQHC in Colorado. Two integrated BHPs provide BH services to almost 4000 patients empaneled to this study's three rural clinics.

Participants

Of the 3962 total registered patients to the three rural health centers, ~90% were Caucasian, 8% Hispanic, and 46% of all patients were at or below 200% of the federal poverty level (\$11 880 for individuals, \$16 020 for a family of two, \$24 300 for a family of four, etc.); 2879 patients were 18 years of age or older, the ultimate pool of participants for this study. A total of 426 adults (14.8%) of this pool completed the PHQ-9 and ORS on presentation to their medical providers. There were 297 women and 129 men with an average age of 46 (age range: 18–82 years, SD = 14.78).

Measures

The PHQ-9 is a nine-item depression scale from the Primary care evaluation of mental disorders (PRIME-MD) diagnostic instrument for common mental disorders, frequently used in primary care (Kroenke *et al.*, 2001). Internal reliability of the PHQ-9 is strong ($\alpha = 0.89$), and a recommended score of 10 or higher has an 88% sensitivity and 88% specificity for major depression (Kroenke and Spitzer, 2002). The PHQ-2 consists of the first two items of the PHQ-9 and for our investigation, we used a PHQ-2 clinical cut-off score of 3 or greater, which has an 83% sensitivity and 90% specificity for major depression (Kroenke and Spitzer, 2002).

The ORS (Miller and Duncan, 2000) assesses four dimensions: (1) individual – personal or symptomatic distress or well-being, (2) interpersonal – relational or

Outcome Rating Scale (ORS)

Name _____ Age (Yrs): ____ Sex: M / F
 Session # ____ Date: _____
 Who is filling out this form? Please check one: Self _____ Other _____
 If other, what is your relationship to this person? _____

Looking back over the last week, including today, help us understand how you have been feeling by rating how well you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels. *If you are filling out this form for another person, please fill out according to how you think he or she is doing.*

Individually
 (Personal well-being)

I-----I

Interpersonally
 (Family, close relationships)

I-----I

Socially
 (Work, school, friendships)

I-----I

Overall
 (General sense of well-being)

I-----I

The Heart and Soul of Change Project

<https://heartandsoulofchange.com>

© 2000, Scott D. Miller and Barry L. Duncan

Figure 1 The Outcome Rating Scale. Source: Reprinted with permission. For examination only. Download a free working copy at <http://heartandsoulofchange.com> or pcoms.com

family distress, (3) social – the patient’s social role functioning, that is, work/school and non-familial relationships, and (4) overall – a big picture perspective or general sense of well-being (see Figure 1). These four dimensions are translated into a visual analogue format of four 10-cm lines where patients place a mark on each line with low scores to the left and high to the right. The score is the summation of the marks made by the patient to the nearest millimeter on each of the four lines, measured

by a centimeter ruler, template, or web system. On the basis of over 400 000 administrations of the ORS and confirming earlier calculations (Miller *et al.*, 2003), Duncan (2014) reported the clinical cut-off for adults as a total score of 25. Adults scoring under 25 are reporting distress typical of individuals receiving psychotherapy, psychotropic medication, or both, and those scoring above 25 are scoring typical of persons who are not receiving treatment. Rated at a Flesch–Kincaid Grade

Level 5 and translated into 24 languages, the ORS is easily understood by patients from a variety of different cultures and has immediate connectivity to a patient's life functioning (Duncan, 2012).

Multiple validation studies of the ORS (Miller *et al.*, 2003; Bringham *et al.*, 2006; Campbell and Hemsley, 2009; Reese *et al.*, 2012) as well as efficacy studies have found that the ORS generates reliable scores. Coefficient α s have ranged from 0.87 to 0.91 in validation studies and from 0.82 (individual therapy) (Reese *et al.*, 2009) to 0.92 (group therapy) (Slone *et al.*, 2015) in clinical studies. Concurrent validity of the ORS has found moderately strong correlations with other validated measures (Miller *et al.*, 2003; Bringham *et al.*, 2006; Campbell and Hemsley, 2009; Gillaspay and Murphy, 2011).

Procedure

The PHQ-9 and the ORS were completed by adult patients on a double-sided sheet when they presented for their primary care health appointment. The measures were introduced, administered, and scored by either medical assistants upon rooming the patient for their medical vitals or front desk staff in the waiting rooms. The measures were only given when an integrated BHP was in clinic so that such screening would provide appropriate BHP back-up and not interfere with workflow demands. Positive screens on the PHQ-9 (total score of 10 or greater) and ORS (total score <25) were reported by the medical assistants to the medical provider, who then had the option of consulting the BHP.

Results

Mean scores for all patients on the ORS ($M = 26.79$, $SD = 10.02$), PHQ-9 ($M = 6.66$, $SD = 6.19$), and PHQ-2 ($M = 1.43$, $SD = 1.69$) were below the respective clinical cut-offs. Coefficient α s for scores on the ORS, PHQ-9, and PHQ-2 were 0.92, 0.89, and 0.81, respectively. Bivariate correlations between the ORS and the PHQ-9 and PHQ-2 were 0.72 and 0.70, respectively. Both coefficients offer evidence of concurrent validity for the ORS. We evaluated the number of patients who were classified in the clinical range on each of the instruments. There was moderate agreement between the ORS and PHQ-9

Table 1 Descriptive statistics and percentage of patients in clinical range for the Outcome Rating Scale (ORS), Patient Health Questionnaire-9 and -2 items (PHQ-9 and PHQ-2)

Measures	M (SD)	% in clinical range
ORS	15.90 (5.69)	39.2 ($n = 167$)**
PHQ-9	14.78 (3.94)	29.6 ($n = 126$)
PHQ-2	4.06 (1.07)	23.2 ($n = 99$)

** < 0.001.

according to $\kappa = 0.56$ ($P < 0.001$), 95% confidence interval (CI 0.48, 0.64); the percentage of agreement was 78.64. There was also moderate agreement between the ORS and PHQ-2, $\kappa = 0.48$ ($P < 0.001$), 95% CI (0.40, 0.56); percentage of agreement was 77. We also conducted a McNemar test given that we had paired nominal-level data to compute if the proportion of patients who scored in the clinical range differed on the two measures. The ORS categorized significantly more patients in the clinical range than either the PHQ-9 χ^2 ($df = 1$, $n = 426$) = 19.78, $P < 0.001$ or the PHQ-2 χ^2 ($df = 1$, $n = 426$) = 47.18, $P < 0.001$ (see Table 1).

Discussion

This preliminary study compared well-validated primary care depression screens (PHQ-9; PHQ-2) with an ultra-brief, four-item global measure of distress across major life domains (ORS) within three family practice, FQHCs. The ORS had never before been investigated in primary care as a universal screener and this investigation explored its capacity to do so in comparison with the PHQ-9. The ORS had robust correlations with the PHQ-9 and PHQ-2, comparable internal consistency, and categorized patients similarly overall. In addition, the ORS classified significantly more patients in the clinical range for potential BH consultation. Although preliminary, these results suggest that an ultra-brief measure of distress across life functioning that also covers the whole developmental age spectrum (Duncan *et al.*, 2006) may cast a wider net and offer a viable alternative to the limitations of traditional symptom-based and diagnostic-specific primary care BH screeners. While we believe engaging more patients in BH intervention to be a positive step to improve

patient outcomes, there may be drawbacks including more demand for BHPs and additional workflow concerns.

A possible concern is the internal consistency estimate of 0.92 for the ORS, indicating potential redundancy (Steiner, 2003). This is likely, in part, due to the high correlation between the last item 'overall' and the first item 'individually' (Campbell and Hemsley, 2009). Although this indicates psychometric redundancy, we believe this concern is mitigated given the inclusion of the last item was for clinical purposes (Duncan, 2012) and reflects a balance between being psychometrically sound and clinically useful.

There are several limitations to this exploratory investigation. Although the ORS demonstrated initial evidence of concurrent validity as indicated by the strong correlation coefficients, other aspects of validity compared with the PHQ-9 were not measured, nor was the ORS's sensitivity and specificity tested. This will be addressed in a follow-up study. Another weakness of this study was that it did not systematically address feasibility (ie, number of patients not screened, number refused, reasons for refusal, impact on clinical schedule or staff workload), nor did we collect data on the number of BH consults triggered by a positive ORS or PHQ-9 and their follow-up outcomes. This too will be addressed in a follow-up study. We believe, however, that the four-item ORS strikes a feasibility balance between the PHQ-2 alone and either a stepped assessment or the PHQ-9 alone. A third weakness was that the patient sample was composed of primarily rural white, female, low-income adults, and our findings may not generalize to other populations. Lastly, the screening measures were not universally nor randomly administered, possibly affecting the study's results.

The ORS, part of the SAMHSA designated evidence-based practice for psychotherapy, PCOMS, may also offer integrated BH care a feasible outcome measure for short-term BH treatment. While identifying patients with psychosocial distress impacting their health and well-being is an important function of primary care screening tools, their use as a quality improvement intervention has not been demonstrated. The PHQ-9, for example, has been shown to be a valid tool for monitoring clinical change over time (Löwe *et al.*, 2004), but has not been empirically demonstrated to improve patient outcomes (Gilbody *et al.*,

2008; Fuchs *et al.*, 2015). The PCOMS feedback intervention has been demonstrated to improve patient outcomes in five randomized clinical trials (Anker *et al.*, 2009; Reese *et al.*, 2009; 2010; Shuman *et al.*, 2015; Slone *et al.*, 2015). Only future research can determine whether these benefits extend to primary care BH intervention.

Acknowledgements

The authors would like to express their appreciation to Peak Vista Community Health Centers, its medical leadership, and the clinic support team of the Health Centers involved in this pilot study.

References

- Anker, M.G., Duncan, B.L. and Sparks, J.A. 2009: Using client feedback to improve couples therapy outcomes: a randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology* 77, 693–704.
- Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K., and Hatcher, S. 2010: Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine* 8, 348–53.
- Bland, R. 2004: Depression and its management in primary care. *Canadian Journal of Psychiatry* 52, 75–76.
- Blasinsky, M., Goldman, H.H. and Unützer, J. 2006: Project IMPACT: a report on barriers and facilitators to sustainability. *Administration and Policy in Mental Health* 33, 718–29.
- Bringhurst, D.L., Watson, C.W., Miller, S.D. and Duncan, B.L. 2006: The reliability and validity of the Outcome Rating Scale: a replication study of a brief clinical measure. *Journal of Brief Therapy* 5, 23–30.
- Campbell, A. and Hemsley, S. 2009: Outcome Rating Scale and Session Rating Scale in psychological practice: clinical utility of ultra-brief measures. *Clinical Psychologist* 13, 1–9.
- Duncan, B., Sparks, J., Miller, S., Bohanske, R. and Claud, D. 2006: Giving youth a voice: a preliminary study of the reliability and validity of a brief outcome measure for children. *Journal of Brief Therapy* 5, 5–22.
- Duncan, B.L. 2012: The Partners for Change Outcome Management System (PCOMS): the Heart and Soul of Change Project. *Canadian Psychology* 53, 93–104.
- Duncan, B.L. 2014. *On becoming a better therapist: evidence-based practice one client at a time*, second edition. Washington, DC: American Psychological Association.
- Duncan, B.L. and Reese, R.J. 2015: The Partners for Change Outcome Management System (PCOMS): revisiting the client's frame of reference. *Psychotherapy* 52, 391–401.

- Edlund, M.J., Unützer, J. and Wells, K.B.** 2004: Clinician screening and treatment of alcohol, drug, and mental problems in primary care: results from healthcare for communities. *Medical Care* 42, 1158–166.
- Fuchs, C.H., Haradvala, N., Hubley, S., Nash, J.M., Keller, M.B., Ashley, D., Weisberg, R.B. and Uebelacker, L.A.** 2015: Physician actions following a positive PHQ-2: implications for the implementation of depression screening in family medicine practices. *Families, Systems, and Health* 33, 18–27.
- Gilbody, S., Sheldon, T. and House, A.** 2008: Screening and case finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal* 178, 997–1003.
- Gillaspay, J.A. and Murphy, J.J.** 2011: The use of ultra-brief client feedback tools in SFBT. In Franklin, C.W., Trepper, T., McCollum, E. and Gingerich, W., editors *Solution-focused brief therapy*. New York: Oxford University Press, 73–93.
- Hunter, C.L., Goodie, J.L., Oordt, M.S. and Dobbmeyer, A.C.** 2009: *Integrated behavioral health in primary care: step-by-step guidance for assessment and intervention*. Washington, DC: American Psychological Association.
- Kroenke, K. and Spitzer, R.L.** 2002: The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals* 32, 1–7.
- Kroenke, K., Spitzer, R.L. and Williams, J.B.** 2001: The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 16, 606–13.
- Kroenke, K., Spitzer, R.L. and Williams, J.B.** 2003: The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care* 41, 1284–292.
- Löwe, B., Kroenke, K., Herzog, W. and Gräfe, K.** 2004: Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders* 81, 61–66.
- Miller, S.D. and Duncan, B.** 2000: *The Outcome Rating Scale*. Jensen Beach, FL: Author.
- Miller, S.D., Duncan, B.L., Brown, J., Sparks, J. and Claud, D.** 2003: The Outcome Rating Scale: a preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy* 2, 91–100.
- Petterson, S., Miller, B.F., Payne-Murphy, J.C. and Phillips, R.L. Jr.** 2014: Mental health treatment in the primary care setting: patterns and pathways. *Families, Systems, and Health* 32, 157–66.
- Reese, R.J., Norsworthy, L.A. and Rowlands, S.R.** 2009: Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy* 46, 418–31.
- Reese, R.J., Toland, M.D. and Kodet, J.** 2012: *Validity of a psychotherapy outcome measure: the Outcome Rating Scale*. Poster presented at Annual Meeting of the American Psychological Association, August, Orlando, FL.
- Reese, R.S., Toland, M.D., Slone, N.C. and Nosworthy, L.A.** 2010: Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy* 47, 616–30.
- Shuman, D.L., Slone, N.C., Reese, R.J. and Duncan, B.L.** 2015: Using client feedback to improve outcomes in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research* 25, 396–407.
- Siu, A.L., US Preventive Services Task Force** 2016: Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. *Journal of the American Medical Association* 315, 380–87.
- Slone, N.C., Reese, R.J., Mathews-Duvall, S. and Kodet, J.** 2015: Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics Theory Research and Practice* 19, 122–36.
- Steiner, D.L.** 2003: Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment* 80, 99–103.
- US Preventive Services Task Force** 2002: Screening for depression: recommendations and rationale. *Annals of Internal Medicine* 136, 760–64.
- US Preventive Services Task Force** 2009: Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 151, 784–92.