The Wellness Assessment:
Global Distress and Indicators of Clinical Severity

May 2010

Background

Research has shown that the integration of outcomes measurement into clinical practice is associated with better clinical outcomes (Herman, Chan, & Zazzoli, 2006; Lambert, 2007; Lambert, Whipple, & Hawkins, 2003; Nierenberg, Ostacher, & Borrelli, 2006). To help ensure appropriate treatment and improve patient outcomes, Optum invested in the development of an outcomes measurement system called ALERT® (***AL****gorithms for* ***E****ffective* ***R****eporting and* ***T****reatment*). This program involves the use of self-report outcomes data that provide objective, patient-centered information.

As an integral part of ALERT, the Wellness Assessment (WA) is a psychometrically-tested instrument for identifying and monitoring change in psychological distress, identifying chemical dependency risk and medical comorbidity, and measuring workplace functioning. Network clinicians are asked to administer the WA at the first outpatient visit with their patients, and again between the third and fifth sessions. Clinicians are free to administer the WA more often if they wish during treatment. Clinicians review the completed WA and submit it to Optum. Four months later, a follow-up WA is mailed to the member by Optum[1]. Algorithms based on the patient WA responses and claims data assist in identifying members at risk. When an at-risk case is identified, an ALERT letter is sent to the clinician highlighting clinical concerns and/or the case is triaged to an Optum Care Advocate for review and outreach.

Initially developed in 1999, the WA was patterned after other well-validated public domain instruments, including subscales from measures such as the Symptom Checklist-90 (SCL-90R; Brophy, Norvell, & Kiluk, 1988) and the Short Form Health Survey (SF-36; Ware, Gandek, & IQOLA Project Group, 1994). Its use in a managed behavioral health organization (MBHO) was tested in a NIMH-funded study (# 1 R43MH57614-01A1) on the effects of administering patient assessments and delivering feedback reports to clinicians. Results from the parent study have been reported elsewhere (Azocar, Cuffel, McCulloch, et al., 2007; Brodey, Cuffel, McCulloch et al., 2005). The WA has undergone a number of psychometric analyses both internally and by an external academically-affiliated third-party. These have been done to ensure its psychometric integrity

[1] Mailing is dependent on having a valid mailing address to which Optum was granted permission to mail by the member.

as an outcome tool and to affirm its use as an objective assessment tool with external credibility.

Global Distress (GD) is the core scale of the WA that encompasses symptoms of anxiety and depression, perceptions of self-efficacy, and functional impairments. The current version of the GD scale was developed following a detailed psychometric evaluation conducted in late 2006 by an external psychometrician (Doucette, 2006). Single parameter Item Response Theory (IRT/Rasch) analyses were used to determine whether the scale should undergo modification. As a result of multiple analyses[2], the scale was shortened and response options were simplified. Using a community sample, the clinical threshold that differentiated a clinical population from a non-clinical population was also defined. In 2008, IRT/Rasch analyses were repeated on a sample of 99,319 WA responses to confirm that the modifications had been effective in improving the scale's psychometric properties and that the scale was sensitive to measuring change over time (Doucette, 2008). The analyses also confirmed that the GD scale displayed good internal reliability (Table 1).

Table 1. Reliability Estimates for Global Distress Scale

Scale Reliability Total Sample	0.87
Measurement Model Reliability	0.90
Cronbach Alpha (Classical Test Theory)	0.90

Optum conducted a preliminary analysis in 2008 on a sample of 32,906 patients with baseline WAs to better understand the relationship between GD and psychiatric diagnoses from administrative claims data. Results demonstrated that greater levels of GD severity were associated with higher diagnosis rates for Mood Disorders (Major Depressive Disorder especially), while the reverse was true for Adjustment Disorders, which were diagnosed more often among respondents with lower levels of GD.

Objective of Current Analyses

The objective of this white paper is to describe analyses that further explored the relationships between psychological distress as measured by the GD scale of the WA and commonly used indicators of clinical severity. This is the first of a series of planned white papers on ALERT and the validity of the WA as an objective assessment tool. The analyses reported herein update and expand upon those

[2] IRT/Rasch analyses included item characteristic curves, item fit and information statistics, person-Item distributions, response scale differentiation/segmentation, Item factors/dimensionality, scale reliability, item correlations, and item frequencies and distributions.

conducted in 2008, and were intended as exploratory analyses to begin to understand the validity of the GD scale in particular.

- Using a more recent sample of WA responses administered under Optum's ALERT program, GD baseline scores were compared to diagnoses from behavioral health claims. The most common disorders among Optum outpatients are Mood Disorders (primarily Depressive Disorders), Adjustment Disorders, and Anxiety Disorders. While GD is not intended to be used as a diagnostic tool, but as a measure of psychological distress among outpatients, we expected to find moderate correlations between GD severity and the presence of these disorders. Specifically, we expected to find that the likelihood of Depressive or Anxiety Disorders increases with GD severity while the likelihood of an Adjustment Disorder diagnosis is greater among patients displaying lower levels of GD severity.
- The same sample was used to compare GD scores to other patient self-report measures on the WA, including workplace impairment, health status, and substance use risk. The intent was to investigate the relationship between GD as a construct and other measures of functioning and clinical risk because we hypothesized there would be some degree of overlap across domains. If the GD scale is a valid measure of distress, we would expect to find associations between GD severity and other measures of impairment on the WA.
- A second sample of data gathered on adults enrolled in Optum's LifeSolutions program (an integrated medical-behavioral outreach support program) was used to explore the relationship between the GD scale and the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001), a widely-used depression measure. While a more extensive cross-validation study of the WA using a wide array of validated measures is planned in partnership with George Washington University for late 2010, this data was available in the interim and allowed us to conduct preliminary analyses. Unlike the PHQ-9, the GD scale is not intended for use as a diagnostic tool for depression. Nevertheless, we believe depressive symptoms may contribute to the construct of global distress, and thus expected to find a moderate correlation between scores on the GD and the PHQ-9.

Method

Sample

Baseline WAs administered by clinicians under ALERT between January 1, 2009 and June 30, 2009 were extracted from the Optum data warehouse. Inclusion criteria for the analyses were WAs with complete GD scales (i.e., no more than

three of the fifteen items missing) that had been administered to the patient by the clinician at the first or second session (i.e., the start of a treatment episode). Additionally, the patient had to be continuously eligible for behavioral health benefits for a 120-day study period (30 days before the baseline WA was completed and 90 days after) and have at least one paid date of service within that study period. If a patient completed a WA that corresponded to the start of more than one treatment episode within this study period, only data pertaining to the earliest episode for the patient was retained in the sample.

Applying the above criteria, the primary sample dataset ("ALERT sample") used in these analyses included WA responses from 40,689 WA patients. In addition, a second, independent sample of cases ("LifeSolutions sample") was identified that consisted of 10,763 patients who completed WA's and PHQ-9s (described below) between April, 2007 and May, 2010 during their participation in LifeSolutions (described above). These assessments were completed at program intake and had to be administered within 7 days of one another. For a respondent's data to be considered complete, no more than three items on the GD scale and no items on the PHQ-9 could be missing for each respondent.

Measures

The 15-item GD scale for adults, as described above, encompasses symptoms of anxiety and depression, perceptions of self-efficacy, and functional impairments. GD scores range from 0 to 45 (higher scores indicating greater severity) and are categorized into severity levels (Table 2).

Table 2. Adult Global Distress Severity Level Descriptions

Total Score	Severity Level
0-11	Low
12-24	Moderate
25-38	Severe
39-45	Very Severe

Other self-report measures of clinical severity and impairment were included in the analysis. Workplace impairment was measured by two items on the WA. Respondents were first asked how many days they were unable to work due to their physical or mental health in the past 30 days (absenteeism). The second item asked how many days respondents had to cut back on how much they could get done due to their physical or mental health (presenteeism). Unemployed respondents were asked to skip these items. General health status in the WA was measured by an item derived from the SF-36 asking respondents to rate their general health. Respondents who rated their health as 'Fair' or 'Poor' were identified as having poor health. Having a co-morbid medical condition was defined by patient self-report indicating the presence of a serious or chronic medical condition. Substance use risk was measured in the WA by three items

from the CAGE-AID scale. Respondents who endorsed at least two items were deemed to be at risk for a Substance Use Disorder.

The PHQ-9 is an instrument specifically designed for use in primary care, but has also been widely used in clinical research. The PHQ-9 contains items that explicitly measure diagnostic criteria for Depressive Disorders from the DSM-IV. PHQ-9 scores range from 0 to 27 (higher scores indicating greater severity), and diagnostic algorithms can be applied to the responses to identify either Major Depressive or Other Depressive Disorders (Spitzer, Kroenke, Williams, & PHQPC Study Group, 1999).

Administrative specialty behavioral health claims data for the ALERT sample were used to identify diagnoses. The most frequently assigned primary diagnosis for the services incurred during the study period was assigned as the patient's diagnosis for the purpose of these analyses.

Results

Sample Demographics

As summarized in Table 3, the majority of respondents were female (65.2%) with a mean age of 39.9 years. The mean GD score was 19.3, equating to moderate severity. Mood, Adjustment, or Anxiety Disorders accounted for 92% of the population.

Table 3. ALERT Sample Demographics (n = 40,689)

Demographics		
Mean Age	39.9 yrs	
	N	**%**
Female	26,525	65.2%
Region		
West	9,482	23.3%
South	11,423	28.1%
Midwest	10,451	25.7%
Northeast	9,333	22.9%
Wellness Assessment Measures		
Mean Global Distress Score	19.3	
Global Distress Severity	**N**	**%**
Low	9,963	24.5%
Moderate	18,425	45.3%
Severe	11,288	27.7%
Very Severe	1,013	2.5%
Mean Days Absent	2.2	
Mean Presenteeism Days	2.4	
Good Health	32,757	80.5%
Medical Co-morbidity	16,920	41.6%

Substance Use Risk	476	11.7%
Claim-based Measures		
Diagnoses	**N**	**%**
Mood Disorders	16,195	39.8%
Adjustment Disorders	14,497	35.6%
Anxiety Disorders	6,871	16.9%
Other Psychiatric Disorders	3,126	7.7%

Respondents in the LifeSolutions sample were similar to respondents in the ALERT sample with respect to GD severity (M=18.6). The average PHQ-9 depression severity level (M=10.5) represented moderate depression, and met criteria for a PHQ-9 diagnosis of either Major or Other Depressive Disorder in 49.5% of respondents. Sixty-nine percent of this sample was female and had a mean age of 51.4 years.

Global Distress and Association with Diagnosis

The prevalence of diagnosis (captured from behavioral health claims) by GD severity level for the ALERT sample is presented in Table 4. Due to the high incidence of Mood Disorders among outpatients and the range of severity contained within the Mood Disorder spectrum of diagnoses, the associations between GD and Major Depressive, Other Depressive (Dysthymia and Depression NOS), and Other Mood Disorders were examined (Table 4). The percentage of patients with Major Depressive and Other Mood Disorders increased with GD severity. The percentage of patients with Adjustment Disorders and Other Psychiatric Diagnoses, on the other hand, decreased with GD severity. Other Depressive Disorders showed a slightly higher incidence among respondents with moderate levels of GD and a low incidence among those with Very Severe GD. The proportion of patients with Anxiety Disorders, however, was relatively consistent across all levels of GD.

Table 4. Rates of Diagnosis by Baseline Global Distress Severity Level

Global Distress Severity Level	N	Percent by Primary Claims Diagnosis					
		Mood Disorder			Adjustment	Anxiety	Other
		Major Depressive	Other Depressive	Other Mood			
Low	9,963	7.9 %	9.4 %	4.7 %	50.6 %	14.7 %	12.7 %
Moderate	18,425	16.6 %	13.0 %	8.7 %	37.2 %	17.9 %	6.6 %
Severe	11,288	29.2 %	11.0 %	15.6 %	21.8 %	17.2 %	5.3 %
Very Severe	1,013	35.2 %	4.6 %	24.9 %	13.9 %	16.6 %	4.7 %

GD scores showed significant, positive correlations with Mood Disorder (r = .28, p<.0001), specifically with Major Depressive Disorder (r = .16, p<.0001) and Other

Mood Disorder (r = .23, p<.0001), and significant, negative correlations with Adjustment Disorder and Other Psychiatric Disorder (r = -.24 and -.11, respectively; p<.0001). There was no correlation between GD score and rates of Anxiety Disorders or Other Depressive Disorders (r = .02 and .00 respectively).

The LifeSolutions sample permitted an additional analysis of the relationship between GD scores and a depression diagnosis derived from the self-report PHQ-9 using algorithms defined by Spitzer et al. (1999). Using the PHQ-9-derived diagnosis, a stronger relationship with GD was found than with the claims-based diagnoses in the ALERT sample. In this sample, GD severity was strongly correlated with the presence of a Major Depressive Disorder (r =.62, p<.0001). Like the claims-based diagnostic results above, there did not appear to be a link between GD and the presence of an Other Depressive Disorder (r =.07).

To further understand the degree of correspondence between GD scores and diagnoses, scores on the GD above the clinical threshold (> 11) were compared with the presence or absence of Depressive, Anxiety, and Adjustment Disorders (Table 5) in order to provide the following metrics:

- *Sensitivity* – the proportion of patients with a given diagnosis who have a clinical GD score.
- *Specificity* – the proportion of patients without a given diagnosis who have a non-clinical GD score.
- *Positive Predictive Value (PPV)* – the proportion of patients with a clinical GD score who are identified with a given diagnosis.
- *Negative Predictive Value (NPV)* – the proportion of patients with a non-clinical GD score who are identified without a given diagnosis.

As shown in Table 5, sensitivity was highest for the claims-based and PHQ-9-based diagnosis of Major Depressive Disorder (89.5% and 99.3% respectively) and the lowest for Other Psychiatric Disorder (59.6%). Specificity was highest for claims-based Mood Disorders in general (31.7%), and for the PHQ-9 diagnosis of Major Depressive Disorder (37.1%). Studies have documented sensitivity and specificity with Major Depressive Disorder as high as 95% for the PHQ-9, and 74% and 81% for the Mental Component Summary scale of the SF-36 (Ware, Gandek, & IQOLA Project Group, 1994). Since the GD scale was not designed to be a diagnostic instrument, but instead intended to measure more global constructs representing depression, anxiety, self-efficacy, and functional impairments, it is not surprising to see lower levels of sensitivity using the clinical threshold of the GD scale than has been reported for the PHQ-9, a tool specifically developed to aid in identifying depressive disorders.

Table 5. Operating Characteristics for Clinical Global Distress & Diagnostic Categories

	Prevalence†	Sensitivity	Specificity	PPV	NPV
Mood Disorders	16,195 (39.8%)	86.5	31.7	45.6	78.0
Major Depressive	7,485(18.4%)	89.5	27.6	21.8	92.1
Other Depressive	4,622 (11.4%)	79.7	25.0	12.0	90.6
Other Mood	4,088 (10.1%)	88.6	25.9	11.8	95.3
Adjustment Disorders	14,497 (35.6%)	65.2	18.8	30.8	49.4
Anxiety Disorders	6,871 (16.9%)	78.7	25.1	17.6	85.3
Other Psychiatric Disorders	3,126 (7.7%)	59.6	23.6	6.1	87.3
PHQ-9 Major Depressive	3,354 (31.2%)	99.3	37.1	41.7	99.2
PHQ-9 Other Depressive	1,979 (18.4%)	91.9	29.7	22.3	94.2

† Prevalence percentages for claims-based diagnoses are based on n = 40,689. For PHQ-9 diagnoses n = 10,763.

Global Distress and Association to Other Patient Self-Report Measures

GD severity was correlated with the other self-report measures of impairment and clinical severity on the WA (Table 6). Both the average workdays impacted by absenteeism and presenteeism as well as the incidence of absenteeism and presenteeism increased with greater GD severity. GD severity was positively related to patients' perceptions of health status and to the likelihood of poor health. The percentage of respondents endorsing medical co-morbidity and substance use risk increased with greater GD severity as well, but were positively correlated to a lesser extent. As depicted in Table 7, there was also a strong positive relationship between the PHQ-9 and GD severity scores (r = .83, p<.0001) among the LifeSolutions sample.

Table 6. Wellness Assessment Self-Report Measures by Global Distress Severity

	Global Distress Severity				
	Low	Moderate	Severe	Very Severe	*r*
Absenteeism					
Mean Days Missed	0.9	1.8	3.7	7.6	.23***
Percent Reporting Absence	17.6%	33.3%	52.9%	69.6%	.30
Presenteeism					
Mean Days Impacted	0.7	2.0	4.5	7.9	.30***
Percent Reporting Impact	11.6%	28.5%	44.6%	53.1%	.29
Health					
Percent Reporting Poor Health	7.0%	15.9%	33.0%	57.8%	.30***
Percent Reporting Medical Co-morbidity	31.6%	40.3%	50.6%	63.1%	.17***
Substance Use					

Percent Endorsing Risk	7.6%	11.8%	14.8%	16.4%	.09***

*** $p < .0001$.

Table 7. PHQ-9 Severity by Global Distress Severity

PHQ-9 Severity	N	Percent by Global Distress Severity			
		Low	Moderate	Severe	Very Severe
None	2,377	83.6 %	16.0 %	0.4 %	0.04 %
Mild depression	2,770	24.5 %	69.3 %	6.2 %	0.04 %
Moderate depression	2,497	3.6 %	68.1 %	28.1 %	0.1 %
Moderately severe depression	19,04	0.7 %	40.3 %	57.6 %	1.4 %
Severe depression	1,215	0.3 %	12.6 %	76.7 %	10.5 %

Summary

The GD scale of the WA was designed to measure general levels of psychological distress common to the population of patients in psychotherapy. These analyses compared GD scores on the baseline WA to clinical severity indicators based on self-reported depression, workplace productivity, overall health, and substance use, as well as clinical severity indicators derived from administrative claims data. The results provide initial support for the validity of the GD scale as a global measure of clinical severity by demonstrating its relationships with other measures of psychiatric acuity and impairment, as well as its ability to appropriately differentiate populations at varying levels along the severity spectrum.

A key indicator of clinical severity is the diagnosis assigned by the treating clinicians. The most prevalent diagnoses among Optum outpatients are Mood Disorders (primarily Depressive Disorders), Adjustment Disorders, and Anxiety Disorders. A higher proportion of patients with Major Depression was found among those with higher levels of GD, while the proportion of patients with Adjustment Disorders declined as GD severity increased. Adjustment Disorders are characterized by less severe symptoms and impairment compared to Mood Disorders; hence, these findings suggest that the GD scale appropriately differentiates clinical severity within these diagnostic categories. There was not, however, an association found between GD severity and the rates of Anxiety Disorders among the respondents. This finding may be an artifact of the diagnostic category of anxiety, given that the disorder encompasses a number of different diagnoses (from phobias to generalized anxiety and panic disorder) with different presentations. To further investigate the association

between GD and Anxiety Disorders, an anxiety scale will be included in the battery of external measures used in the cross-validation study planned for late 2010.

Patients with depression diagnoses accounted for 30% of the ALERT sample and 50% of the LifeSolutions sample. The analysis suggested that GD severity was most closely related to the rates of Major Depressive Disorder (regardless of being diagnosed by the clinician or derived from patient self-report on the PHQ-9), rather than Other Depressive Disorders (Dysthymia and Depression NOS). While this analysis revealed a weaker association between GD severity and the diagnosis of Other Depressive Disorders, there was a strong correlation between GD severity and severity of depression as measured by the PHQ-9 in the LifeSolutions sample. These correlations may reflect the fact that both the GD and PHQ-9 are self-report measures, whereas the diagnoses are based on a nosology used by clinicians to interpret and categorize symptomatology. Taken together, these findings again underscore that while the GD scale correlates with depression severity, it should not be used as a diagnostic tool.

The ALERT sample included other patient self-report measures, including indicators of workplace impairment, health status, and substance use risk. While these scales measure distinct domains, we would expect to see linkages to GD severity. The analysis partially supported these expectations: the rates and intensity of workplace impairment and the presence of health concerns were moderately correlated with GD severity, as was, to a lesser extent, the presence of medical co-morbidity and substance use risk.

Although these analyses highlighted relationships between psychological distress as measured by the GD scale of the WA and commonly used indicators of clinical severity, future work will include evaluating the criterion validity of the GD scale against other gold standard measures of psychological distress, depression, anxiety, and workplace productivity.

References

Azocar F., Cuffel, B.J., McCulloch, J., McCabe, J., Tani, S., & Brody, B. (2007). Monitoring patient improvement and its relation to treatment outcomes in a Managed Behavioral Health Organization. *Journal of Healthcare Quality, 29*, 2, 4 -- 12.

Brodey, B.B., Cuffel, B., McCulloch, J., Tani, S., Maruish, M., Brodey, I., Unutzer, J. (2005). The acceptability and effectiveness of patient-reported assessments and feedback in a managed behavioral healthcare setting. *Am J Manag Care, 11*, 774 -- 780.

Brophy, C.J., Norvell, N.K., & Kiluk, D.J. (1988). An examination of the factor structure and convergent and discriminant validity of the SCL-90R in an outpatient clinic population. *Journal of Personality Assessment, 52*, 334 – 340.

Doucette, A. (2006) Selected findings: Adult Wellness Assessment. *Unpublished Technical Report*.

Doucette, A. (2008). Selected findings: Adult Wellness Assessment: Second pilot study Adult Wellness Assessment. *Unpublished Technical Report*.

Herman, R., Chan, J., Zazzali, D. (2006). Aligning measurement-based quality improvement with implementation of evidence-based practices. *Administration and Policy in Mental Health and Mental Health Services Research,* 33:636-645..

Kroenke, K., Spitzer, R.L, & Williams, J.B. (2001). The PHQ-9: Validity of a brief depression measure. *Journal of General Internal Medicine, 16*, 606 – 613.

Lambert, M. (2007). Presidential address: What have we learned from a decade of research aimed at improving psychotherapy outcomes. *Psychotherapy Research,* 17 1-14.

Lambert, M., Whipple, J., Hawkins, E,, Vermeersch, D.A., Nielsen, S.L, & Smart, D.W. (2003). Is it time for clinicians to track patient outcomes on a routine basis?: A meta-analysis. *Clinical Psychology: Science & Practice, 10*, 288 -- 301.

Nierenberg, A.A., Ostacher, M.J., Borrelli, D.J., Iosifescu, D.V., Perlis, R.H., Desrosiers, A., Armistead, M.S., Calkins, A.W., & Sachs, G.S. The integration of measurement and management for the treatment of bipolar

disorder: a STEP-BD model of collaborative care in psychiatry. *Journal of Clinical Psychiatry, 67*, Suppl 11, 3 -- 7.

Spitzer, R. L., Kroenke, K., Williams, J.B., & PHQPC Study Group. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA, 282*, 1737 – 1744.

Ware, J.E., Gandek, B., & IQOLA Project Group. (1994). The SF-36 Health Survey: Development and use in mental health research and the IQOLA Project. *International Journal of Mental Health, 23*, 49-74.