

SELECTED PSYCHOMETRIC PROPERTIES
OF THE OUTCOME QUESTIONNAIRE-45
IN AN OLDER ADULT POPULATION

by

Kevan LaFrance

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Educational Psychology

The University of Utah

May 2013

Copyright © Kevan LaFrance 2013

All Rights Reserved

ABSTRACT

The Outcome Questionnaire 45 (OQ-45), a 45-item self-report measure of psychological distress, was examined for internal consistency, test-retest reliability, concurrent validity, and construct validity in a sample of adults age 60 years and older. Sixty-six community volunteers recruited from local educational and community service workshops participated in the study that included filling out the OQ-45. Of these 66 volunteers, 34 (57%) completed a second OQ-45 by mail approximately three weeks after the first administration. Measures of depression (Geriatric Depression Scale), anxiety (Geriatric Anxiety Inventory) and general functioning (Short Form-12) were obtained at the first administration and OQ-45 scores were correlated with these measures. The OQ-45 internal consistency estimates were consistent with estimates found in younger adults, though some problematic items were identified. Test-retest analysis showed scores remained relatively stable between administrations. An existing database of 611 older persons who had received treatment in a community mental health setting was used as a contrast to the 66 healthy volunteers. OQ-45 scores from the volunteer community were significantly lower than those collected from the existing database. Implications for these findings for clinical use and for future research are discussed, including recommendations for improving the design of the OQ-45 for older adult groups.

For my wife, who made this achievement possible and worthwhile.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapters	
1 REVIEW OF THE LITERATURE	1
The Outcome Questionnaire-45	1
Characteristics of an Older Population	11
Problem Statement and Hypotheses	15
2 METHODS	19
Subjects	19
Instruments	22
3 RESULTS	24
Descriptive Results	24
Hypothesis 1	25
Hypothesis 2	26
Hypothesis 3	27
Hypothesis 4	28
Hypothesis 5	29
4 DISCUSSION	42
Internal Consistency and Problematic Items	43
Test-Retest Reliability	46
Concurrent and Construct Validity	49
Limitations of This Study	51
Recommendations for Future Research	53
Recommendations for Clinical Use	55
REFERENCES	59

LIST OF TABLES

Table	Page
1. Community and Clinical Sample Descriptive Statistics	33
2. Community Sample ($N = 66$) Low Item-Total Correlations	34
3. Clinical Sample ($N = 611$) Low Item-Total Correlations	34
4. Community Sample ($N = 66$) Frequencies of Missing Values by OQ-45 Item . . .	35
5. Clinical Sample ($N = 611$) Frequencies of Missing Values by OQ-45 Item	36
6. Community Sample Descriptive Statistics for Missing Items 7 and 17.	37
7. Correlation Coefficients between OQ-45 and Measures of Mental Health and Physical Health Symptoms in the Community Sample	38
8. Community Sample ($N = 66$) and Published Research Data with Younger Adults: Comparison of Concurrent Validity Estimates of the Outcome Questionnaire-45	39
9. Comparison of OQ-45 Scores by Community and Clinical Groups with Published OQ-45 Data	40
10. Sensitivity and Specificity of Calculated Cutoff Score of 54 on the OQ-45	41

LIST OF FIGURES

Figure	Page
3.1. Histogram of OQ-45 scores in a community sample	31
3.2. Histogram of OQ-45 scores in a clinical sample	32
3.3. Test-retest reliability estimate of OQ-45 scores in the community sample	34

CHAPTER 1

REVIEW OF THE LITERATURE

The Outcome Questionnaire-45

Psychological Distress and Instrument Development

The conceptualization, assessment, and the impact of therapy on change in indices of psychological distress is a prominent theme in the psychotherapy outcome literature. Self-report symptom questionnaires have been developed and standardized to gauge psychological distress primarily in terms of its severity and duration. These measures, for the most part, have identified specific domains of symptomatology that encompass emotional, somatic, and relational issues that are commonly linked to distress.

Among the self-report instruments used to measure distress, the Outcome Questionnaire-45 is one of the most prominent (OQ-45). The OQ-45 was developed in the early 1990s to track change in distress over time (Lambert et al., 2004) and it has frequently been employed in psychotherapy outcome studies to measure change in distress as a consequence of therapy. The OQ-45 is a self-report measure consisting of 45 Likert-type scale items generated for three primary content domains: symptom distress, problems in interpersonal relationships, and problems related to social role performance (Lambert et al., 1996). Each item is scored on a 5-point scale from 0-4 (0 = “Never” to 4 = “Almost Always”), with positive items (i.e., “I enjoy my spare time”) reverse-scored. A lower OQ-45 domain/total score indicates low distress. OQ-45 total

scores range from 0 to 150 with a total score of 64 as a “cutoff” for distress that would warrant psychological treatment; that is, a total score at or greater than 64 indicates clinically significant distress (Lambert et al., 2004). A deviation interval of 14 points in OQ-45 total score indicates clinically significant change (Jacobsen and Truax, 1991; Lambert et al., 2004). Mean scores for the OQ-45 range from ($M = 45.19$; $SD = 18.57$) in a non-clinical community group (age data unavailable) and from ($M = 42.15$; $SD = 16.61$) to ($M = 51.34$; $SD = 24.45$) in healthy student groups. In clinical and distressed student groups, OQ-45 scores have been reported to range from ($M = 73.61$; $SD = 21.39$) to ($M = 88.80$; $SD = 26.66$; Lambert et al., 2004).

Overview of the OQ-45

The OQ-45 total score is purported to measure a general mental health factor that Lambert et al. (2004) have labeled as psychological distress (de Jong et al., 2007; Lambert, 2007; Mueller, Lambert, & Burlingame, 1998). In fact, the first domain of the OQ-45 is defined as symptom distress. This domain is comprised of 25 items that specifically reflect symptoms associated with diagnosable mental disorders that, according to de Jong et al. (2007) are linked to expressed distressed mood (for example: “I have difficulty concentrating;” “I feel fearful;” “I feel worthless”). The second or interpersonal problems scale and consists of items that reflect the presence of interpersonal distress, perceptions of personal inadequacy, and feelings of isolation (for example, “I feel loved and wanted;” “I am satisfied with my relationships with others;” “I have trouble getting along with friends and close acquaintances”). The third scale, social role performance, measures distress in fulfilling social roles in work, school, family, and leisure activities (for example, “I enjoy my spare time;” “I feel angry enough at

work/school to do something I might regret,” “I feel that I am doing well at work/school”).

As a general framework for item completion, the respondent is instructed to answer the questions in response to the following time-referenced statement:

“Looking back over the past week, including today, help us understand how you have been feeling.”

The 45 items follow this overview statement. Thus the OQ-45 is designed to measure the respondent’s general psychological state over the past week, as opposed to measuring how a person feels at any specific moment in time.

The OQ-45 has been used to inform clinicians about client distress and to assess client therapy progress in a variety of settings including university counseling centers and community mental health agencies (Lambert et al., 2004). The following psychometric review of validity and reliability studies of the OQ-45 addresses the question of whether: (1) it adequately measures the constructs it was designed to measure and (2) if OQ-45 scores are stable over time.

OQ-45 Psychometric Review

Construct Validity of the OQ-45

The OQ-45 has been subject to substantial empirical scrutiny to evaluate if specific items conform to the conceptual domains around which the OQ-45 was constructed; namely, symptom distress, interpersonal problems, and social role performance. An initial report of a confirmatory factor analysis reported by Mueller et al. (1998), found that OQ-45 items did not cluster around these domains. Rather, the 45 items conformed to a single factor which the authors subsequently labeled psychological

distress. The failure to identify three distinct factors in the OQ-45 has been attributed to high inter-item correlations among the 45 items. The single factor solution supports other research that has consistently reported very high internal consistency coefficients among the OQ-45 items (Lambert et al., 2004). Despite the lack of empirical support for the three conceptual subscales of the OQ-45, Mueller et al. (1998) have argued that a total OQ-45 score retains its originally intended utility as a meaningful measure of psychological distress. Therefore, the OQ-45 total score has been viewed as an acceptable measure of therapy outcomes to assess therapy efficacy (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009; Minami et al., 2009) and to monitor client progress during treatment (Finch, Lambert, & Schaalje, 2001; Whipple et al., 2003).

Concurrent Validity

The concurrent validity of the OQ-45 has been examined in several studies that have found it to have been highly correlated with other instruments purported to measure psychopathology including the Symptom Checklist-90-Revised (SCL-90-R; $r = .78$; Lambert et al., 1996; Beckstead et al., 2003; de Jong et al., 2007). Lambert et al. (2004) found that the OQ-45 total score was highly correlated with the Beck Depression Inventory (BDI; $r = .80$), Zung Self Rating Depression Scale (ZSDS; $r = .88$), the Zung Self Rating Anxiety Scale (ZSAS; $r = .81$), the Taylor Manifest Anxiety Scale (TMA; $r = .86$), and the SF-36 Medical Outcome Questionnaire ($r = .81$). In addition it was moderately correlated with the Inventory of Interpersonal Problems (IIP, $r = .54$) and the Social Adjustment Scale (SAS, $r = .65$; 2004) and it has been found to be positively associated with number of psychiatric diagnoses as measured by the Structured Clinical Interview for DSM Disorders (SCID) ($r = .87$; Lueck, 2004). These concurrent validity

estimates support the OQ-45 as a measure of psychological distress primarily reflected through mental health symptomology. A defining feature of the OQ-45 is that its purpose is to measure general psychological distress, whereas most of these other instruments are designed for more specific elements of psychopathology such as depression (BDI, ZSDS), anxiety (ZSAS, TMA), and social functioning (IIP, SAS).

Doerfler, Addis, and Moran (2002) analyzed the convergent and divergent validity of the OQ-45 with the Behavior and Symptom Identification Scale (BASIS-32). The BASIS-32 is a brief clinician-administered instrument constructed to measure symptom and behavioral distress in psychiatric inpatient populations (Eisen, Dill, & Grob, 1994). The BASIS-32 is comprised of five subscales; namely, psychosis, impulsive and addictive behavior, relation to self and others, depression and anxiety, and daily living and social role functioning. Doerfler et al. (2002) found a moderate correlation between OQ-45 total score and BASIS-32 total score ($r = .64$). Doerfler et al. concluded that the OQ-45 and the BASIS-32 are both equally useful in measuring client progress and therapy outcomes, although the BASIS-32 may be better suited for an inpatient psychiatric population. BASIS-32 items are specifically worded to reflect symptoms and concerns commonly reported by psychiatric inpatients, whereas the OQ-45 items reflect more general symptoms and experiences of a broader range of respondents.

Reliability of the OQ-45

Lambert et al. (2004) reported that the OQ-45 total score is an internally consistent and stable measure of distress. In support of this contention, Lambert et al. (1996) collected data from a variety of adult populations including two groups of persons

that were construed as a “normative” nonclinical sample (college students and a general community sample) as well as three different clinical adult groups (University outpatient clinic participants, community mental health center outpatient clients, and individuals from an Employee Assistance Program who were seeking mental health treatment).

Lambert et al. (2004) found that in these groups there was high internal consistency across items ($\alpha = 0.93$, $N_{Student} = 157$; $\alpha = 0.93$, $N_{Clinical} = 289$) and high test-retest reliability in the student group ($r = 0.84$; 3-week retest interval; Lambert, 1996).

Normative Data

The normative data collected to date for the OQ-45 has been primarily for persons in the age range of young to middle-aged adult populations (between 18 and 59 years of age; Lambert et al., 2004). With respect to persons within this age grouping, Lambert et al. (2004) identified a diverse community sample through the Utah County phone directory as well as by identifying persons at random who were listed as insured by a large national insurance firm of over 800 employees from various businesses in Ohio. This initial normative sample of Utah community residents and insured employees in Ohio consisted of 815 participants who reported that they had never received a mental health diagnosis. In addition to this community sample, Lambert et al. recruited 538 undergraduates from classroom settings in Utah, Idaho, and Ohio to represent a college-age normative group. Lambert et al. also created a clinical group utilizing data from a sample of 441 individuals seeking mental health treatment through an Employee Assistance Program (EAP). Lambert et al. summarized age grouping in these clinical and community samples and found no significant differences between age groups. However, of note for the current study, of the 504 participants in an EAP sample, only eight of

these persons were over the age of 60 years. In these normative samples, no gender differences were found.

The OQ-45 has been administered to a wide variety of racial and ethnic groups and, as noted previously, it has been translated and normed across a number of non-English speaking groups including participants whose first language was Spanish, Italian, Dutch, Norwegian, Japanese, Hebrew, Arabic, Russian, German, French, and Swedish (Lambert et al., 2004). Similar to the results found with English-speaking samples from the United States, high internal consistency estimates were found in a Dutch sample ($\alpha = .92$; $n = 2623$; de Jong et al., 2007), and in an Italian sample ($\alpha = .92$; $n = 823$; Lo Coco et al., 2008). High internal consistency estimates ($\alpha = .94$; $N = 398$) were also obtained from a sample of 398 adults who resided in Utah and whose primary language was Spanish (Jurado, 2007).

Some studies have found differences in mean OQ-45 scores between different demographic groups (Gregersen, Nebeker, Seely, & Lambert, 2004; Lambert et al., 2004) and that cutoff scores vary, especially for non-English speaking groups. Lambert et al. (2004) reported that such variations more likely reflected cultural differences in response patterns rather than indicating that one group may be more distressed than another.

Regardless of the characteristics of the recruited group, however, a change in score over time has been supported as indicative of change in psychological distress during treatment (Lambert et al., 2006). Lambert et al. (2004) report that while some studies may suggest differences in mean scores between various racial and ethnic groups, repeated administrations and test-retest reliability evaluation of the OQ-45 within these groups supported the contention that OQ-45 scores are a reliable measure of

psychological distress across a wide variety of persons and is a valid measure of change in distress symptomatology. Of interest in the current study is that the reliability of the OQ-45 has not yet been examined in older adult groups, that is among persons who are 60 years or older.

Critique of the OQ-45

As noted previously, the OQ-45 has three apparent strengths: (1) the relative ease of administration which facilitates measuring progress in treatment. (2) high internal consistency and concurrent validity estimates as reported in a variety of studies with multiple populations and across persons who speak different languages, and (3) its accessibility to clinicians and treatment centers around the globe. As a measure of change, the OQ-45 is one of the more widely used instruments to document psychotherapy outcomes in the published literature. However, the OQ-45 also has some limitations that require further examination, particularly the lack of empirical support for the theorized three-factor domain structure, a test-retest artifact (namely, respondents appear to spontaneously report diminished distress over time), and the lack of empirical examination of reliability and validity with older persons.

A first limitation of the OQ-45 is the problematic and relatively weak empirical support for the theorized three-domain factor structure. As indicated previously, Mueller et al. (1998) found no statistical support of a three-factor model in an American sample. Interestingly, the three-factor solution was supported in a Dutch sample (de Jong et al., 2007) and a four-factor model was found in an Italian sample (Lo Coco et al., 2008). Thus, it appears that the factor structure of the OQ-45 may vary depending on the cultural background of the sample under investigation. While a consistent factor structure

between samples has not emerged in studies with the OQ-45, it has been argued that the OQ-45 still retains its value as a global measure of psychological distress (Lambert et al., 2004).

A second limitation of the OQ-45 relates to the test-retest reliability of the instrument. An artifact of repeated administrations of the OQ-45 is a tendency for respondents to report reduced distress with repeated testing regardless of participation in treatment (Vermeersch, Lambert, & Burlingame, 2000). Vermeersch et al. (2000) found that in a nontherapy control group, OQ-45 score values (or levels of psychological distress) decreased over time. This finding could not be accounted for as regression toward the mean. In other words, the reduction in scores between administrations was not found solely in respondents with extreme initial scores. This test-retest artifact or tendency to report fewer and/or less severe symptoms over repeated administrations of psychological tests has been found in previous research with other psychological measures (Ahava, Iannone, Grebstein, & Shirling, 1998; Durham et al., 2002). Durham et al. (2002) summarize the previous findings and attempt to explain this artifact as the product of social desirability (attempting to present oneself in a positive light). Thus it would appear that the OQ-45 may be vulnerable to the some of the same reliability and validity limitations reported for other similar self-report measures, such as the Beck Depression Inventory (Ahava et al., 1998).

However, despite bias towards this systematic change in OQ-45 score over time for nontherapy controls, Vermeersch et al. (2000) reported that the participants who were receiving therapy during the study demonstrated significantly greater reduction in OQ-45 scores across repeated testing intervals than a control group. However, it is also possible

that the effect of social desirability could be greater within a clinical sample; that is, persons in treatment may try to please the therapist by reporting “improvement”.

Durham et al. (2002) evaluated the test-retest artifact in the OQ-45 and found this effect to be less than one third of the Reliable Change Index of 14, noting that OQ-45 scores were observed to generally drop significantly between first and second administrations and demonstrated no reliable change with further administrations. Durham et al. asserted that the test-retest artifact was not affected by frequency of administration and also determined that only .7% of the variation in scores was attributable to social desirability. Thus, while the artifact of spontaneous improvement in OQ-45 score between first and second administration is statistically significant with sufficiently large sample sizes, this pattern of scores decreasing over time does not reach clinical significance. This underscores the importance of proper interpretation of scores utilizing approved metrics of change (Lambert et al., 2004).

A third limitation, and perhaps the most significant weakness of the OQ-45 is the lack of empirical support for use in groups of persons to which it is routinely administered, namely adults who are 60 years and older (Roseborough, Luptak, McLeod, & Bradshaw, 2013). There is no normative data on OQ-45 performance on persons who are 60 years and older. Older adults have unique characteristics that differ from younger adults and these differences should be evaluated before routinely employing the OQ-45 in older adult clients. For example, many older adults are no longer actively engaged in work or school which creates problems in the interpretation of OQ-45 items that highlight these terms. This necessitates the gathering and analysis of normative data for older

persons. To lay the foundation for such an evaluation, the following literature review will highlight some of the characteristics and unique concerns of older adults.

Characteristics of an Older Population

Physical and psychosocial challenges increase in older people with advancing age. This issue is a prominent public health concern as the average life expectancy of persons in our Western society has increased to nearly 80 years. These physical and psychological concerns impact a person's perceptions of health and wellbeing. Three domains where wellbeing diminishment has been documented in older adults include: (1) physical health issues associated with age-related decline, (2) restricted social networks, and (3) mental health symptoms that are interrelated with physical and social functioning.

Physical Health

As people advance in age, they experience an increase in health conditions (Hill, 2005; Hill, 2011). Physical health concerns are more prevalent among the elderly, with estimates that approximately 62% of adults over age 65 have multiple chronic medical conditions (Vogeli et al., 2007). Health conditions prevalent among older adults include arthritis, hypertension, heart disease, cancer, diabetes, and stroke (Hill, 2005). These health conditions can contribute to limited mobility, increased likelihood of injury or accidents, and reduced ability to engage in self-care. Such difficulties contribute to increased psychological distress and mental health symptoms and decreased quality of life (Hill, 2005; Vink, Aartsen, & Schoevers, 2008).

Social Networks

Many seniors are also faced with social restrictions not often encountered by a general younger adult population. As individuals age, voluntary social networks (such as friendships and greater community engagement) often shrink. As a consequence, older adults tend to intensify their focus on intimate partnerships and close family relationships (American Psychological Association [APA], 2004). This narrowing or selectivity of social networks could be attributed to a developmental shift in focus to more emotionally satisfying and deeper relationships and experiences (Carstensen, Isaacowitz, & Charles, 1999). The shrinking of social networks may also be due to functional limitations and other life changes which restrict one's ability to access social resources and support (Edelstein, Kalish, Drozdick, & McKee, 1999). Physical health problems that impact an older person's day-to-day interaction with others can place stress on intimate relationships and create tension in close affiliative relationships with children or other family members who may have different ideas about how much help is expected for the older individual in need of support (APA, 2004). The size of social networks may also be limited by the death of friends or family members, which in turn creates restrictions in family structure (Hill, Thorn, & Packard, 2000). In particular, older adults are more likely to experience the death of a spouse than younger adults (Hill, 2005). This can be particularly detrimental as social support systems frequently provide a buffer against distress (Rogers & Delewski, 2004). Long-term marriage may positively impact chronic health problems, functional limitations, and disability (Pienta, Hayward, & Jenkins, 2000), but loss of these relationships when they are most needed in old age can also be a source of substantial trauma. Thus, older adults who encounter loss in social

relationships can be impacted negatively on perceived health and well-being measures (APA, 2004). These varying factors—developmental changes, physical health and functional limitations, death and loss—result in older adults in general having a different and more mixed experience of social networks than is typically found among younger adults (Stephens, Alpass, Towers, & Stevenson, 2011).

Mental Health

Older adults experience unique mental health challenges as a function of advancing age including difficulties in physical health and social networks that could result in negative mood or distress. The epidemiological data estimate a prevalence of chronic mental illness in the elderly population at approximately 26% of persons 65 years and older (Rogers & Delewski, 2004), and suicide rates are higher among older adults than in younger age groups (APA, 2004). From a developmental perspective, distress experienced by older adults may be attributed at times to the specific challenges of navigating developmental tasks associated with aging, such as dealing with existential realities associated with death and loss including physical declines, social limitations, or complicated forms of grief (APA, 2004).

An example of the interaction of physical, social, and mental health functioning is provided by Gurland, Wilder, and Berkman (1988), who note that perceived distress resulting from physical disability is intensified by attitudes of hopelessness and helplessness which cause older persons to further limit their social networks (i.e., supportive friends and family). They also pointed out in their review that when older adults experience grief and bereavement due to loss of friends and relatives to death, they are at risk for experiencing diminished social support, which can in turn contribute to

mental health symptoms such as increased sadness, depression, and distress.

Furthermore, in a more recent comprehensive overview of research comparing risk factors for depression and anxiety in older adults, Vink, Aartsen, and Schoevers (2008) found a number of risk factors affecting incidence and prevalence of depression, which included poor self-perceived health, functional disability, decreased social network size, and chronic disease. Conversely, mental health difficulties such as intense life regrets have also been found to be predictive of increased health problems over the life span (Wrosch, Dunne, Scheier, & Shulz, 2006). This body of research suggests that the mental health challenges of older adults could be different from those experienced by younger adults.

Conclusion

In summary, psychological distress in later life is related to a number of different factors among older adults including age-related changes in physical health, social functioning, and mental health concerns, all of which contribute to increased psychological distress. To effectively apply treatments or interventions to assist an older adult clinical population, accurate assessment of psychological distress within this population is essential (Alexopoulos, 2005; APA, 2004; Yon & Scogin, 2007). Gallo & Bogner (2006) underscore the importance of accurate assessment of the psychological and functional needs of older adults. They suggest that even small improvements in psychological health in older persons can have substantial positive effects on perceived quality of life. Additionally, the APA (2004) "Guidelines for Psychological Practice with Older Adults" specifically states the necessity of utilizing assessments that have been shown to be valid and reliable with older adults and as has been noted in previous

research, a special set of empirically based treatments have been developed for older adults (Scogin, 2007). However, the issue of accurate assessment of mental health symptoms becomes challenging given that many instruments that focus on distress have not been developed with the older adult in mind. Thus more research is needed to ascertain if instruments created to assess distress in a younger adult population (such as the OQ-45) may be applicable to the elderly or, if not, how such instruments may be adapted for this population. The lack of empirical research examining the use of the OQ-45 with a geriatric population was the impetus for the current study which investigated reliability and validity issues related to uses of the OQ-45 in persons 60 years and older.

Problem Statement and Hypotheses

There are no reports in the published literature that evaluate the psychometric properties of the OQ-45 in an older adult population. However, the OQ-45 is commonly employed to assess psychological state in clinical settings where geriatric clients (those over 60 years of age) receive mental health services (Roseborough et al., 2013). Despite its lack of empirical support, placement and treatment decisions are made based on OQ-45 scores. As was noted in the literature review, although older adults present similar mental health and physical concerns as younger aged persons, there are notable differences in the experience and presentation of mental health concerns between younger and older persons.

Hypothesis 1: The reliability estimates of internal consistency of OQ-45 items will not be statistically different than internal consistency estimates reported in the younger adult age groups.

Considering the potential role of psychological distress in community mental health programming in older adult groups, and in light of current trends to administer the OQ-45 to increasingly older aged persons, a goal of this study was to examine selected aspects of reliability and validity of the OQ-45 within an older aged sample. As a case in point, some items of the OQ-45 may be construed differently to an older retired person than to a young mid-career adult, such as the items pertaining to work and/or school, which may not be applicable to seniors who are retired. Further, seniors may not always understand how to interpret such items (i.e., “I feel stressed at work/school”). A thorough exploration of the OQ-45 as it applies to older adults is necessary to ensure that the instrument is useful as a measure of general psychological distress among persons in this age group. For this reason, the internal consistency of OQ-45 items and the contribution of each item to overall instrument reliability was assessed as part of the overall analysis plan of this study; that is, an effort was made to examine the stability (test-retest reliability) of the OQ-45 in an older-aged sample.

Hypothesis 2: The test-retest reliability estimate of the OQ-45 with older adults will not be statistically different from test-retest estimates reported in the younger adult age groups.

Hypothesis 3: As with previous research, scores on the OQ-45 will decrease between first and second administrations of the OQ-45, though this difference will be small and will be less than the Reliable Change Index (RCI) of 14 points.

Individuals who present with physical health problems will not demonstrate this decrease.

The test-retest artifact noted in previous research also bears examination in an older adult group. As noted above, previous research has found that samples of persons who were not participating in treatment showed spontaneous improvement in OQ-45 scores between first and second administrations. Durham et al. (2002) found that this improvement was a 4-point drop in total distress score, which is less than 1/3 of the 14-point change considered to be clinically significant. However, many older adults, as discussed above, experience physical ailments and health issues which are progressive over time. Thus, it could be expected in older adults with chronic health issues that OQ-45 scores could reflect greater distress with the passage of time as a consequence of worsening health with advancing age. In other words, it may be that as the prevalence of physical health concerns increases with passage of time in older adults, and this could interact with this previously observed phenomenon of spontaneous decrease in OQ-45 scores with repeated administrations.

Hypothesis 4: The OQ-45 correlates with measures of: (a) depression, (b) anxiety, and (c) general quality of life. Correlations found in younger populations with similar instruments will fall within the 95% confidence Intervals of correlations found with older adults.

OQ-45 validity issues were examined in an older adult sample as part of the study design including concurrent validity and criterion validity. It should be noted that of the various instruments used thus far to examine the concurrent validity of the OQ-45, none have been specifically validated in older adults. However, measures do exist that have evaluated psychological distress in geriatric clients. One such instrument is the Geriatric Depression Scale (GDS; Yesavage et al., 1983). The GDS is a 30-item questionnaire

with a “yes-no” item format. The GDS has been widely reported as a valid instrument for measuring depressive symptomatology in both in- and outpatient older adult samples. It has also been asserted based on empirical data that the GDS is a reliable instrument for use with older adults (Edelstein et al., 1999). Another example of a measure created for older adults is the Geriatric Anxiety Inventory (GAI; Pachana et al., 2007). The GAI is a 20-item questionnaire with an “agree/disagree” format that contains a number of items that focus on psychological distress. Given that there exist in the extant literature instruments purported to be reliable and valid measures of symptom expression that are specific to older adults, these have not been used as validity checks for the OQ-45.

Hypothesis 5: Older adults who are receiving treatment for mental health issues in this sample will score significantly higher on the OQ-45 than older adults found in a general community sample.

This study was conducted in two phases to specifically address this hypothesis. The first phase consisted of gathering normative data for the OQ-45 from a small community sample of older adults with no reported mental health conditions. The second phase of the study involved analyzing extant data from a larger clinical sample of older adults who were receiving treatment for mental health issues.

CHAPTER 2

METHODS

Subjects

Community Sample Subjects

A community sample of 66 healthy ambulatory individuals was recruited from educational seminars made available to members of the American Association of Retired Persons (AARP), and through a training workshop for the Foster Grandparent/Senior Companion program which is a local volunteer program of low income older adults who are engaged in community service through Salt Lake County Aging Services. These 66 persons were from a total available sample of approximately 250 to 300 persons who were present in these two contexts. Participants ranged in age from 58 to 85 years ($M = 70.65$ years). There were 49 females, 15 males, and 2 participants did not indicate their sex (74% female and 23% male). The average level of education of participants was 14.7 years (12 .0 years = a high school graduate). Regarding marital status, 45.5% reported that they were married, 28.8% divorced, 16.7% widowed, 4.5% single, 3% cohabitating, and 1.5% did not answer the relationship status question.

Community Sample Data Collection Procedure

After a brief verbal presentation by the researcher, participants were administered a paper and pencil questionnaire packet at each of the respective workshop (or training)

sessions. The packet included a coversheet with informed consent information; a demographic sheet that included date of administration and questions about age, sex, marital status, and whether the participant was currently in treatment for mental health issues; the OQ-45; the GDS; GAI; and SF-12. Completion time of the questionnaires was approximately 30 minutes. Each packet included a self-addressed and stamped envelope, allowing the participants to take the packets home to complete and then mail them back to the researcher. During the verbal instructions for completing the packet and in the packet cover letter, participants were invited to include their return address so that they could be mailed a follow-up packet consisting of only the OQ-45 and an instruction sheet. Follow-up surveys were mailed approximately 2 weeks after the participant completed and mailed back the first set of questionnaires.

For purposes of privacy and confidentiality, a database matching identification numbers and address information was maintained separately from other data. This database with mailing information was deleted once follow-up data were collected. Subjects were provided an incentive in the form of being entered into a random drawing for five participants to receive a free self-help book at an approximate value of \$15 per book to complete and return the follow-up materials. Of the 66 persons who chose to participate in the study, 34 individuals (52%) returned the follow-up questionnaires.

Clinical Sample Data from Extant Database

In addition to this healthy community group that consisted of recruited volunteers, an extant clinical sample was also obtained from a dataset that was in place at Valley Mental Health (VMH). VMH is a local community mental health treatment agency that is a State of Utah service provider for persons with chronic mental health issues.

Included among these client records were older adults, defined in the dataset as the subgroup of persons who were 60 years of age and older. This clinical database sample included information available from the OQ-45 with completed surveys from 611 individuals who ranged in age from 60 to 91 years ($M = 69.05$ years). Of these, 433 were female (70.9%) and 178 were male (29.1%).

The process of accessing this VMH database involved contacting the treatment center program manager of research and evaluation for Valley Mental Health, who has a PhD in psychology. This individual facilitated the obtaining of requisite permissions from the State of Utah to access the data. Once required permissions were obtained, data extraction began by identifying all client records of persons age 60 years and older who received services at VMH and were administered the OQ-45 for the first time between the years 2009 and 2011. The resulting final dataset was comprised of the following variables for each patient: age, sex, date of first administration of the OQ-45, values for each item on the OQ-45 from the first administration on record, and OQ-45 total score. Three patient records were excluded from analysis due to excessive missing data.

Where intermittent missing values for items on the OQ-45 were present in both samples, missing values were replaced utilizing the individual's appropriate subscale mean score, rounded to the nearest whole number. This is the recommended missing values substitution strategy reported by Lambert et al. (2004). As this method would increase the Cronbach's Alpha coefficient, thus inflating the internal consistency estimates, the internal consistency analysis was completed prior to filling in missing data.

Instruments

Outcome Questionnaire-45 (OQ-45)

The OQ-45 is a 45-item questionnaire utilizing a five-point Likert-type response format ranging from “Never” to “Almost Always.” Estimated completion time of the OQ-45 is approximately 5 minutes.

Geriatric Depression Scale (GDS)

The GDS is a 30-item questionnaire with a “yes/no” response format, completion time of this instrument is approximately 5 minutes. The GDS was normed on older adult groups (Yochim, Lequerica, MacNeill, & Lichtenberg, P.A. 2008) and is widely used as a measure of depressive symptoms in geriatric populations. The internal consistency reliability of the GDS has been estimated at $\alpha = .94$ (LaBuda & Lichtenberg, 1999; Yesavage et al., 1983; Yochim et al., 2008).

Geriatric Anxiety Inventory (GAI)

The GAI is a 20-item questionnaire with an “agree/disagree” response format and expected completion time of 3 to 4 minutes (Pachana et al., 2007). The GAI consists of items developed to measure symptoms of anxiety in elderly people including questions pertaining to fearfulness, anxious mood, and somatic symptoms of anxiety. Pachana et al. (2007) noted that the GAI does not focus on exclusively somatic symptoms as there can be considerable overlap with somatic symptoms and expressed anxiety in older persons. Internal consistency estimates of the GAI have been reported to be $\alpha = .91$ and the GAI has been found to have correlations in the range of .57 to .70 with a variety of other well-established measures of anxiety (Pachana et al., 2007).

Short-Form 12 (SF-12)

The SF-12 is a shortened 12-item version of the SF-36, which is a measure of functional health and well-being (Ware, Kosinski, & Keller, 1996). The 12 items were derived from the SF-36. SF-12 items reflect two component scales: a Physical Component Summary (PCS) and Mental Component Summary (MCS; Ware et al., 1996). These two SF-12 scales have demonstrated test-retest reliability estimates of $r = .89$ and $r = .76$ for the PCS-12 and MCS-12, respectively (Ware et al., 1996). While no internal consistency estimates for the SF-12 have been reported, the developers of the SF-12 and SF-36 argue that the internal consistency estimates for the SF-36 version of the scales (PCS-36; $r = .92$; and MCS-36; $r = .88$) are applicable as the SF-12 items account for over 90% of the reliable variance of the original SF-36 PCS and MCS scale scores (Ware et al., 1996). However, as test length contributes to internal consistency estimates, it can be postulated that analysis of the SF-12 would yield lower internal consistency estimates. The SF-12 was employed in this study due to its shorter length and ease of administration. When interpreting scores from the SF-12, higher scores indicate higher states of wellness and functioning, with lower scores indicating the presence of more problems. The SF-12 data were scored using SF Health Outcomes Scoring Software 3.0.

CHAPTER 3

RESULTS

Descriptive Results

Descriptive statistics for both community and clinical samples are presented in Table 1. Distributions of OQ-45 total scores for the community and clinical samples are presented in Figure 1 and Figure 2, respectively.

Community Sample

Mean scores for the 66 community participants on the OQ-45 was 44.76 ($SD = 17.72$). Of these persons, 65 completed the additional measures, with mean scores as follows: GDS ($M = 5.86, SD = 5.17$), GAI ($M = 2.20, SD = 3.49$), and SF-12 PCS ($M = 45.22, SD = 11.49$) and MCS ($M = 53.43, SD = 7.76$). An exploratory correlation analysis revealed no statistically significant correlations between age and any of these test scores.

Clinical Sample

The older adult clinical database sample consisted of 611 individuals with an average age of 69.05 years ($SD = 6.25$). Sex distribution in this sample was 71% female and 29% male. Mean scores for the single OQ-45 administration for this group was 71.83 ($SD = 28.66$).

Hypothesis 1

Hypothesis 1: The reliability estimates of internal consistency of OQ-45 items will not be statistically different than internal consistency estimates reported in the younger adult age groups.

Internal consistency analysis (Cronbach's Alpha) in the community sample ($\alpha_{community} = .92$; $N = 66$) and in the clinical sample ($\alpha_{clinical} = .94$; $N = 611$) were compared to the published results from previous research ($\alpha_{Community} = 0.93$, $N = 157$; $\alpha_{Clinical} = 0.93$, $N = 289$; Lambert et al., 2004). These internal consistency estimates were compared to the previous published research utilizing the Fisher *r*-to-*Z* transformation, subtracting the difference in *Z* scores, and then dividing the difference by the pooled variance. This yielded a $Z_{difference}$ of .45 ($N = 66$; $p > .05$) for the community sample and a $Z_{difference}$ of 1.04 ($N = 611$; $p > .05$) for the clinical sample, which supports the hypothesis.

Items were examined for item-total correlations. This analysis revealed six items in the community sample with low item-total correlations (Table 2) and four such items in the clinical sample (Table 3). Notably, the four items identified as having low item-total correlations in the clinical sample were included in the set of six items in the community sample. Three of the problematic items identified among both samples were related to drinking and/or drug use (OQ-45 items 11, 26, and 32). These three items had a disproportionate percentage of participants marking "Never" (94% for item 11, 92% for item 26, and 97% for item 32 in the community sample; 90% for item 11, 86% for item 26, and 93% for item 32 within the clinical sample), and the distributions of these items was skewed (skewness of 5.71, 4.05, and 5.61, respectively for the community sample;

3.80, 3.20, and 5.06 for the clinical sample). Items 14 (“I work/study too much”), 27 (“I have an upset stomach”), and 28 (“I am not working/studying as well as I used to”) were not skewed (.11, .42, and .37, respectively in the community sample).

OQ-45 items were also examined for frequency of missing values (left blank by participants). Items with missing values in more than 5% of cases are indicated in Table 4 (Community sample) and Table 5 (Clinical sample). In both the community and clinical samples, two items were frequently left blank. These items were Item 7 (“I feel unhappy in my marriage/significant relationship;” missing in 18.2% of cases in the community sample, 19.3% in the clinical sample) and Item 17 (“I have an unfulfilling sex life;” 16.7% missing in the community sample, 19.1% in the clinical sample). As these items focus on interpersonal relationships, the relationship status (e.g., married, divorced, widowed) of the subset of community participants who left these questions blank was examined (Table 6; relationship status data was not available for the clinical sample). Notably, none of the individuals who left item 7 blank endorsed being married or cohabitating with a romantic partner, and only one person who left item 17 blank endorsed being married.

Hypothesis 2

Hypothesis 2: The test-retest reliability estimate of the OQ-45 with older adults will not be statistically different from test-retest estimates reported in the younger adult age groups.

To examine the stability of OQ-45 scores (i.e., whether scores measure the trait of anxiety and acute depression), correlations between first and second administrations were evaluated in the community group. Of the original 66 community participants, 34 (52%)

completed the follow-up OQ-45. The mean OQ-45 second administration score was 42.59 ($SD = 21.50$), with an average time between first and second administrations being 20.75 days ($SD = 3.93$).

The findings from this analysis yielded a test-retest correlation of $r = .91$ ($p < .05$; $n = 34$; see Figure 3) versus a published test-retest correlation of $r = .84$ ($n = 157$; Lambert et al., 2004). These test-retest correlation estimates were compared utilizing the Fisher r -to- Z transformation, subtracting the difference in Z scores, and then dividing the difference by the pooled variance. This yielded a $Z_{difference}$ of 1.56 ($p > .05$), which was nonsignificant.

Hypothesis 3

Hypothesis 3: As with previous research, scores on the OQ-45 will decrease between first and second administrations of the OQ-45, though this difference will be small and will be less than the Reliable Change Index (RCI) of 14 points.

Individuals who present with physical health problems

will not demonstrate this decrease.

A paired-samples t-test was conducted to ascertain if OQ-45 total scores may have changed systematically vs. randomly between first and second administrations in the community sample. This analysis was used to determine if the same test-retest artifact found in previous research (Vermeersch et al., 2000) was also present in this older adult sample (that scores on the OQ-45 decrease between first and second administrations in nonclinical populations). Consistent with the hypothesis, scores between first ($M_{Time1} = 44.03$; $SD = 19.16$) and second ($M_{Time2} = 42.59$; $SD = 21.50$) administrations of the OQ-

45 in this older adult community sample decreased between administrations ($M_{Time1 - Time2} = 1.44$; $t(33) = .96$; $p > .05$). This decrease, however, was not statistically significant and was not correlated with physical health complaints as indicated by the PCS (Physical Component Scale) of the SF-12 ($r = -.11$; $p > .05$) as hypothesized.

Hypothesis 4

Hypothesis 4: The OQ-45 correlates with measures of: (a) depression, (b) anxiety, and (c) general quality of life. Correlations found in younger populations with similar instruments will fall within the 95% confidence Intervals of correlations found with older adults.

The Geriatric Depression Scale (GDS), Geriatric Anxiety Inventory (GAI), and SF-12 were three instruments with which the concurrent validity of the OQ-45 was examined in this community sample. Correlation coefficients were calculated on the OQ-45 total score and the scores of the GDS and GAI and the mental health scale of the SF-12 (MCS), respectively. The OQ-45 total score was correlated with these measures, with the strongest positive correlation between the OQ-45 and the GDS ($r = .73$; $p < .05$; 95% CI [.59, .83]; see Table 7). The OQ-45 was also positively correlated with the GAI ($r = .57$; 95% CI [.38, .72]; $p < .05$) and negatively correlated with the SF-12 MCS ($r = -.45$; 95% CI [-.61, -.20]; $p < .05$). This negative correlation is due to the fact that among the OQ-45, GDS, and GAI, a high score indicates more problems, where the reverse is the case for the SF-12 scales.

Confidence intervals (95%) for the correlations were calculated and compared to correlations reported in previous research (Lambert et al., 2004; See Table 8). This confidence interval evaluation indicated that the published correlation between the Beck

Depression Inventory (BDI) and the State-Trait Anxiety Inventory-State (STAI-S) scale fell within the 95% confidence interval of the correlations between the OQ-45 and similar instruments (the GDS and GAI, respectively).

Hypothesis 5

Hypothesis 5: Older adults who are receiving treatment for mental health issues in this sample will score significantly higher on the OQ-45 than older adults found in a general community sample.

Mean scores and standard deviations for both the community and clinical samples in this study and data from previous published research with younger adult groups are presented in Table 9. An independent samples t-test was conducted to contrast mean scores of the community and clinical samples. This contrast was statistically significant.

The results indicated that the clinical sample ($M_{Clinical} = 71.83$; $SD = 28.66$) was roughly 27 points higher than the community sample ($M_{Community} = 44.76$; $SD = 17.72$) and this difference was significant ($t(675) = -7.52, p < .01$). However, Levene's test for equality of variances indicated that the variances between samples were not equal ($F = 20.17, p < .01$), which is problematic in that it violates an assumption of a standard t-test (Hayes, 1994). While the statistical software used in this study concurrently produced a t-test with unequal variances assumed ($t(106) = -10.96, p < .01$), the fact that this analysis included unequal sample sizes increases the likelihood of a Type 1 error (Keppel, 1991). To examine the extent to which unequal sample sizes may have affected the results of this study, 66 data points were randomly selected from the clinical sample to create equal sample sizes and the means were again compared ($M_{Community} = 44.76, SD = 17.72$; $M_{Clinical} = 69.44, SD = 27.83$). A subsequent t-test with unequal variances assumed ($F =$

16.58, $p < .01$) indicated that the difference in OQ-45 total score between groups was statistically significant ($t(110) = -6.01, p < .01$).

After determining that OQ-45 scores of clinical and community groups of older adults were statistically different, a sensitivity-specificity analysis was conducted to determine a cutoff score. Sensitivity is the ability of the OQ-45 to identify true positives, or the ability to correctly identify an individual from the clinical sample. Specificity is the ability of the OQ-45 to accurately classify an individual from the community sample as not clinically distressed. A cutoff score to distinguish between the clinical and community samples in this study was computed in accordance with the recommendations provided by Lambert et al. (2004). The following formula was used to calculate the cutoff score:

$$c = \frac{(SD_1)(mean_2) + (SD_2)(mean_1)}{SD_1 + SD_2}$$

Utilizing the recommended formula, a cutoff score of 54 was computed (as opposed to 64 in the general population), which resulted in a Sensitivity of .71 and Specificity of .70, respectively. In other words, a score of 54 or above correctly identified a person with significant clinical symptomology 71% of the time (see Table 10). Utilizing the cutoff score of 64 established with younger adult samples would yield a Sensitivity of .57 and a Specificity of .85. Thus the OQ-45 would correctly identify someone with significant clinical symptomology only 57% of the time with the previously established cutoff score.

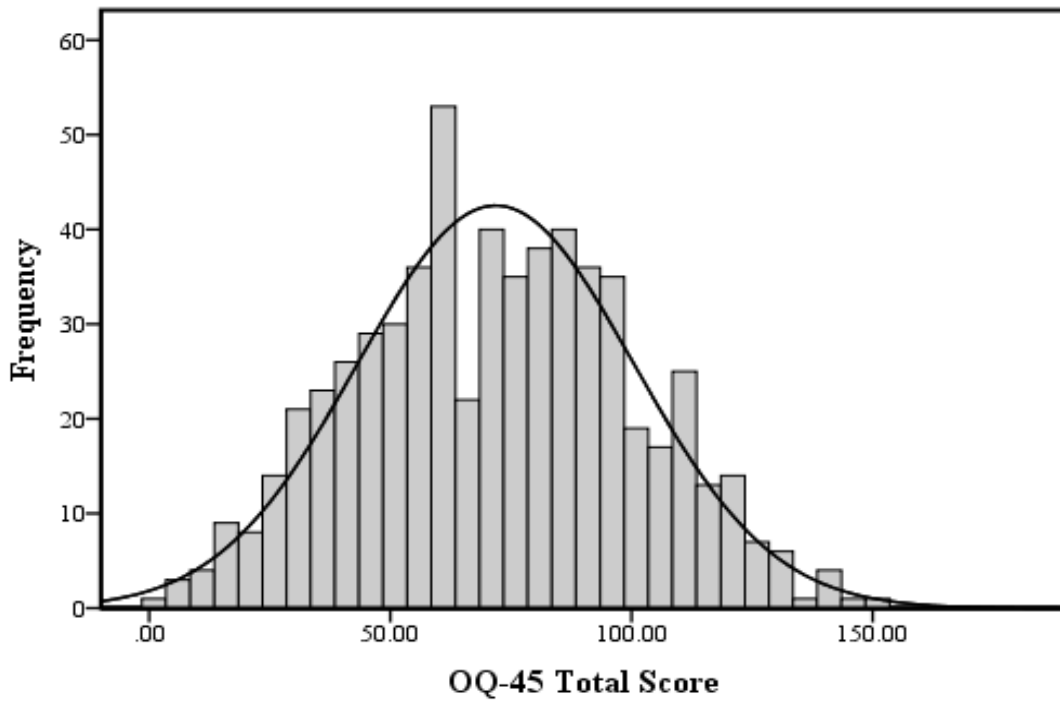


Figure 3.1. Histogram of OQ-45 scores in a Community Sample ($N = 66$).

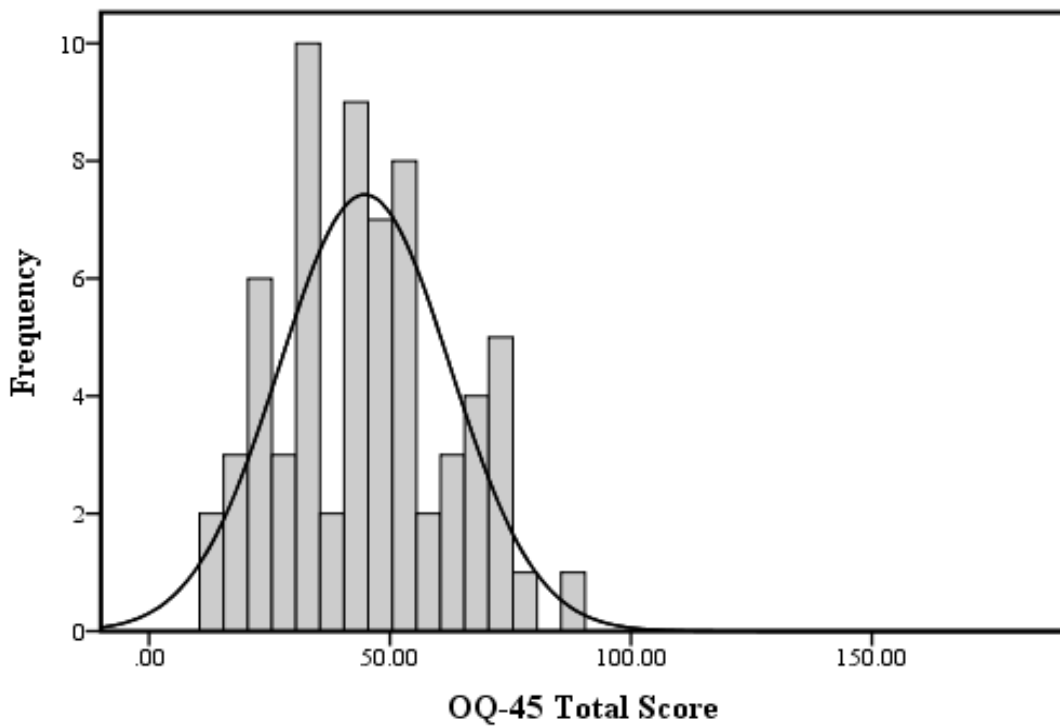


Figure 3.2. Histogram of OQ-45 scores in a Clinical Sample ($N = 611$).

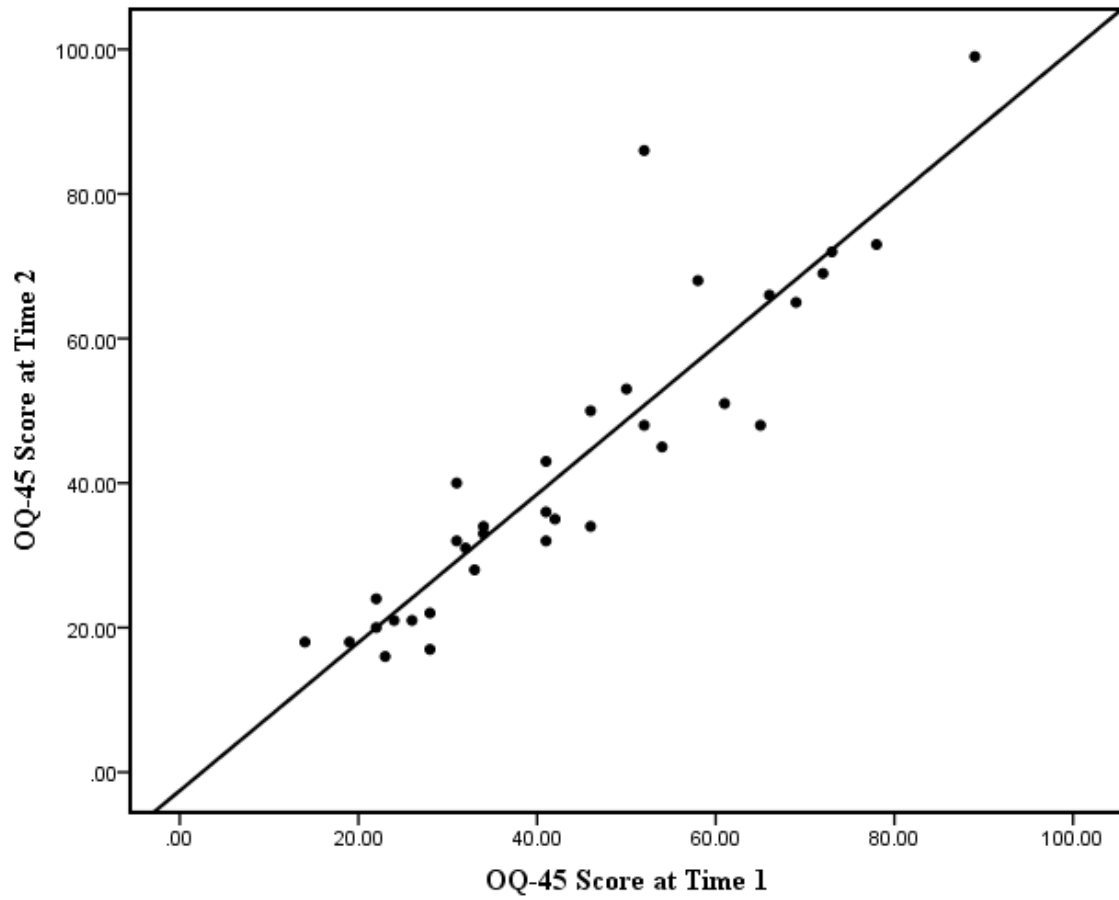


Figure 3.3- Test-retest reliability estimate of OQ-45 scores in the community sample ($r(34) = .91$). The average time between administrations was $M = 20.75$ days.

Table 1

Community and Clinical Sample Descriptive Statistics

Variable	Sample			
	<i>Community (n = 66)</i>		<i>Clinical (n = 611)</i>	
	Range (or <i>n</i>)	Mean (or %)	Range (or <i>n</i>)	Mean (or %)
<i>Age (Years)</i>	Range: 58-85	<i>M</i> = 70.65 (<i>SD</i> = 6.25)	Range: 60-91	<i>M</i> = 69.05 (<i>SD</i> = 7.42)
<i>Sex</i>				
Female	<i>n</i> = 49	74%	<i>n</i> = 433	71%
Male	<i>n</i> = 15	23%	<i>n</i> = 178	29%
Not indicated	<i>n</i> = 2	3%		
<i>Years of Education</i>	Range: 10-21	<i>M</i> = 14.8* (<i>SD</i> = 2.52)		
<i>Marital Status</i>				
Currently married	<i>n</i> = 30	45.5%		
Divorced	<i>n</i> = 19	28.8%		
Widowed	<i>n</i> = 11	16.7%		
Single	<i>n</i> = 3	4.5%		
Cohabiting	<i>n</i> = 2	3%		
Not indicated	<i>n</i> = 1	1.5%		

* 12 years = HS graduate

Table 2

Community Sample (N = 66) Low Item-Total Correlations

OQ-45 Item	Item as stated on OQ-45	Item-Total Correlation
11.	After heavy drinking, I need a drink the next morning to get going. (If you do not drink, mark “never”)	.17
14.	I work/study too much	.09
26.	I feel annoyed by people who criticize my drinking (or drug use) (if not applicable, mark “never”)	.03
27.	I have an upset stomach	-.05
28.	I am not working/studying as well as I used to	.11
32.	I have trouble at work/school because of drinking or drug use (If not applicable, mark “never”)	.06

OQ-45 Cronbach’s Alpha = .92

Table 3

Clinical Sample (N = 611) Low Item-Total Correlations

OQ-45 Item	Item as stated on OQ-45	Item-Total Correlation
11.	After heavy drinking, I need a drink the next morning to get going (If you do not drink, mark “never”)	.10
14.	I work/study too much	.04
26.	I feel annoyed by people who criticize my drinking (or drug use) (if not applicable, mark “never”)	.16
32.	I have trouble at work/school because of drinking or drug use (If not applicable, mark “never”)	.14

OQ-45 Cronbach’s Alpha = .941

Table 4

Community Sample (N = 66) Frequencies of Missing Values by OQ-45 Item

OQ Item	Count Missing (Percent)
7. I feel unhappy in my marriage/significant relationship	12 (18.18%)
17. I have an unfulfilling sex life	11 (16.67%)
34. I have sore muscles	8 (12.12%)
6. I feel irritated	8 (12.12%)
36. I feel nervous	7 (10.61%)
12. I find my work/school satisfying	4 (6.06%)
14. I work/study too much	4 (6.06%)
28. I am not working/studying as well as I used to	4 (6.06%)
40. I feel something is wrong with my mind	4 (6.06%)
44. I feel angry enough at work/school to do something I might regret	4 (6.06%)

Table 5

Clinical Sample (N = 611) Frequencies of Missing Values by OQ-45 Item

OQ Item	Count Missing (Percent)
7. I feel unhappy in my marriage/significant relationship	118 (19.31%)
17. I have an unfulfilling sex life	117 (19.15%)
37. I feel my love relationships are full and complete	51 (8.35%)
12. I find my work/school satisfying	47 (7.69%)
38. I feel I am not doing well at work/school	43 (7.04%)
39. I have too many disagreements at work/school	40 (6.55%)
4. I feel stressed at work/school	39 (6.38%)
28. I am not working/studying as well as I used to	38 (6.22%)
14. I work/study too much	37 (6.06%)

Table 6

*Community Sample Descriptive Statistics for Missing Items 7** and 17****

Variable	Sample			
	Item 7 (n = 12)		Item 17 (n = 11)	
	Range (or n)	Mean (or %)	Range (or n)	Mean (or %)
<i>Age (Years)</i>	Range: 64-85	<i>M</i> = 72.42 (<i>SD</i> = 6.97)	Range: 64-85	<i>M</i> = 73.73 (<i>SD</i> = 6.51)
<i>Sex</i>				
Female	<i>n</i> = 10	83.3%	<i>n</i> = 8	72.7%
Male	<i>n</i> = 0	0%	<i>n</i> = 1	9.1%
Not indicated	<i>n</i> = 2	16.7%	<i>n</i> = 2	18.2%
<i>Years of Education*</i>	Range: 13-20	<i>M</i> = 16.08 (<i>SD</i> = 2.07)	Range: 13-20	<i>M</i> = 15.64 (<i>SD</i> = 2.16)
<i>Marital Status</i>				
Currently married	<i>n</i> = 0	0%	<i>n</i> = 1	9.1%
Divorced	<i>n</i> = 8	66.7%	<i>n</i> = 5	45.5%
Widowed	<i>n</i> = 2	16.7%	<i>n</i> = 3	27.3%
Single	<i>n</i> = 1	8.3%	<i>n</i> = 2	18.2%
Cohabiting	<i>n</i> = 0	0%	<i>n</i> = 0	0%
Not indicated	<i>n</i> = 1	8.3%	<i>n</i> = 0	0%

* 12 years = HS graduate

**OQ-45 Item 7: *I feel unhappy in my marriage/significant relationship.*

***OQ-45 Item 17: *I have an unfulfilling sex life.*

Table 7

Correlation Coefficients between OQ-45 and Measures of Mental Health and Physical Health Symptoms in the Community Sample

Criterion	OQ-45	GDS	GAI	SF-12 MCS
OQ-45 ^a	--	.73**	.57**	-.45**
GDS ^b		--	.70**	-.61**
GAI ^c			--	-.31*
SF-12 MCS ^{***d}				--

$N = 65$; One Community respondent completed only the OQ-45

* Correlation is significant at the $p < .05$ level (2-tailed); $N = 65$

** Correlation is significant at the $p < .01$ level (2-tailed); $N = 65$

*** For PCS and MCS, a lower score indicates more problems, whereas on the OQ-45 and other measures, a lower score indicates fewer problems

^a Outcome Questionnaire-45 (OQ-45)

^b Geriatric Depression Scale (GDS)

^c Geriatric Anxiety Inventory (GAI)

^d Short Form-12 Mental Component Scale (MCS)

Table 8

*Community Sample (N = 66) and Published Research Data¹ with Younger Adults:
Comparison of Concurrent Validity Estimates of the Outcome Questionnaire-45*

Instrument	Correlation with OQ-45 (N = 65)	95% Confidence Interval	Comparable Instrument ¹	Correlation with OQ-45 ¹
GDS ^a	$r = .73$.59 to .83	BDI ^d	$r = .80$ (N = 115)
			ZSDS ^e	$r = .88$ (N = 71)
GAI ^b	$r = .57$.38 to .72	ZSAS ^f	$r = .81$ (N = 71)
			STAI-S ^g	$r = .64$ (N = 115)
			STAI-T ^g	$r = .80$ (N = 115)
MCS ^c	$r = -.45$	-.61 to -.20	SF-36 ^h	$r = .81$

¹From Lambert et al., 2004

^aGeriatric Depression Scale

^bGeriatric Anxiety Inventory

^cShort Form-12 Mental Component Scale

^dBeck Depression Inventory

^eZung Self Rating Depression Scale

^fZung Self Rating Anxiety Scale

^gState Trait Anxiety Inventory (S = State Anxiety; T = Trait Anxiety)

^hShort Form 36 Medical Outcome Questionnaire; sample size not available

Table 9

*Comparison of OQ-45 Scores by Community and Clinical Groups with Published OQ-45 Data**

Sample	Age	N	OQ-45 Mean Score	SD	Range
Community	70.65	66	44.75	17.72	76
Clinical	69.05	611	71.83	28.66	150
Published Data in the General Adult Population					
*Community	N/A	815	45.19	18.57	N/A
*EAP Clinical Services	N/A **	441	73.61	21.39	N/A
*Outpatient Clinics	N/A	342	83.09	22.23	N/A

* From Lambert et al., 2004; Age data from these studies was not available. Subjects for the Community normative study were gathered via selection of every 10th name from the phone book as well as distribution to employees from a large national insurance firm. Subjects for the Employee Assistance Program (EAP) study were gathered from a database supplied by an EAP program in which the OQ-45 was administered as part of the subjects receiving mental health services from the EAP provider. Subjects for the Outpatient Clinics normative study were gathered from a community mental health center in Ohio and subjects were administered the OQ-45 during the normal course of treatment.

** Age of this sample was divided into ranges, with $n = 8$ individuals over age 60. Mean age data were not published.

Table 10

Sensitivity and Specificity of Calculated Cutoff Score of 54 on the OQ-45

Criterion (Clinical vs Nonclinical)	Test Condition	
	OQ-45 < 54	OQ-45 ≥ 54
Clinical (<i>n</i> = 625)	<i>n</i> = 175 False Negative (FN) (Type II Error)	<i>n</i> = 448 True Positive (TP)
Nonclinical (<i>n</i> = 54)	<i>n</i> = 38 True Negative (TN)	<i>n</i> = 16 False Positive (FP) (Type I Error)
	Specificity = .70** TN / (FP + TN) 38 / (16 + 38) = .70	Sensitivity = .72* TP / (TP + FN) 448 / (448 + 175) = .72

* 448 (72%) of Clinical participants were correctly identified as “Clinical”

** 38 (70%) of Nonclinical participants were correctly identified as “Nonclinical”

CHAPTER 4

DISCUSSION

The purpose of this study was to assess aspects of the psychometric properties of the OQ-45 when it is administered to an older adult sample. These analyses included an examination of internal consistency and item review for potentially problematic items, test-retest reliability, concurrent validity, and construct validity. The sample consisted of 66 older adults in the general community. Data for a contrasting clinical sample were also extracted from an existing database containing information on older adults who had received psychiatric diagnoses and who currently were receiving treatment for substantial mental health concerns.

This study found that the OQ-45 has good internal consistency and test-retest reliability, similar to previous published research with this instrument in younger-aged samples. The OQ-45 also demonstrated concurrent and criterion validity in that scores on the OQ-45 were correlated with published measures of anxiety, depression, and general mental health functioning. The OQ-45 was also correlated with a measure of physical health and this correlation was smaller than correlations between OQ-45 total score and mental health measures. Finally, OQ-45 scores were significantly different when contrasting community and clinical samples. Those in the clinical sample expressed higher levels of psychological distress than those in the community sample. The sensitivity and specificity properties of the OQ-45 were assessed using the clinical and a

community sample, respectively. In the next section, the results for the tested hypotheses will be interpreted with respect to published OQ-45 outcomes.

Internal Consistency and Problematic Items

Hypothesis 1: The reliability estimates of internal consistency of OQ-45 items will not be statistically different than internal consistency estimates reported in the younger adult age groups.

When applied to an older adult group, the OQ-45 demonstrated similar internal consistency reliability estimates as when applied to a younger group, which is consistent with the hypothesis. Internal consistency estimates found in previous research were not significantly different from similar estimates found in both community and clinical samples in this study. It should be noted that internal consistency estimates are influenced by number of items (Cortina, 1993) and thus the alpha reported in this study may reflect the sizable number of items that query along a similar content theme; namely, sources of personal distress. The internal consistency estimates ranging from .92 to .94 in this study suggest very limited item-specific variance in the total score of the OQ-45. To better understand how individual items contribute to these estimates, OQ-45 items were analyzed to determine the item-total correlations within both the community and clinical samples. What follows is a discussion of items identified as problematic due to low item-total correlations and items which were frequently left blank by respondents.

Within the community sample, six items were identified as having weak item-total correlations ($r < .20$; see Table 2). Of these items, those pertaining to substance abuse (items 11, 26, and 32) were highly skewed with the majority of respondents indicating *Never*. This could be attributed to cultural factors: namely, that many residents

of the geographical region in which this sample was collected belong to a conservative religious community which abstains from substance use. Religious affiliation was not assessed in this study. Additionally, two of the problematic items were related to work/school (14 and 28) and were left blank more often than other items (see below). The final item in the community sample with a low item-total correlation (item 27; “I have an upset stomach”) was neither skewed nor left blank frequently. Other items related to physical symptoms were not problematic for item-total correlations. It is possible that this particular physiological experience of having an upset stomach does not share the same connection to psychological distress in older adults as in younger adults. The problematic relationship between somatic symptoms and experienced anxiety in older adults was noted by the developers of the Geriatric Anxiety Inventory (Pachana et al., 2007). By contrast, in the clinical sample, only four items (11 “After heavy drinking, I need a drink the next morning to get going,” 14 “I work/study too much,” 26 “I feel annoyed y people who criticize my drinking (or drug use),” and 32 “I have trouble at work/school because of drinking or drug use.”) had item-total correlations less than .20. These items have previously been identified as problematic in that they have low factor loadings, particularly item 14, “I work/study too much” (de Jong et al., 2007; Mueller et al., 1998) and are vulnerable to floor effects (average ratings between *Never* and *Rarely*; Vermeersch et al., 2000).

A small subset of OQ-45 items was judged as problematic in that these items were frequently left blank by respondents. Of particular concern were those items related to work and/or school as well as items that elicited information about romantic relationships and the individual’s sexual behavior. An example of an OQ-45 item in this category was

Item 7, “I feel unhappy in my marriage/significant relationship.” What follows in the next paragraph is an exploration of potential explanations for why these items were left blank.

A possible explanation for non-response to the items pertaining to work and/or school among older adults is that many participants in this sample were retired and/or unable to work, particularly as many of the participants were recruited from a meeting of the American Association of Retired Persons. These work/school items are less likely to be perceived as directly applicable to an older adult’s daily life if the older person is retired and not pursuing further formal education. The questionnaire instructs participants that “work is defined as employment, school, housework, volunteer work, and so forth,” inviting the participants interpret these words flexibly. However, each question includes the explicit words “work/school” or “working/studying” and not alternative words such as “volunteering” or “hobbies.” It is possible many individuals did not thoroughly read the instructions, and thus concluded that the questions about work/school were not applicable to them. It should be noted that most of these items, despite their higher frequency of being left blank, still contributed to the overall high internal consistency of the instrument (as noted previously, the internal consistency estimates were calculated prior to filling in missing values). The exception to this is item 14 (“I work/study too much”).

Items relating to intimate relationships were judged to be problematic as they were also frequently left blank. Like the work/school items, this could be due to the general instructions of the instrument, which provided limited alternatives for flexibly interpreting the terms and ideas presented in the questions. In this sample, 50% of

respondents indicated being widowed, divorced, or single, possibly many of these persons may be uncertain as to how to respond to a query about their current relationship status. While the instructions on the OQ-45 provide general ideas for subjects on how to construe work/school questions (i.e., “work” includes housework, volunteering, etc.), as well as how to interpret questions about drug or alcohol use (i.e., mark “never” if the question does not apply), neither the instructions nor the questions themselves instruct an individual how to specifically interpret words that may have multiple meanings for an older adult. One example is Question 17 (“I have an unfulfilling sex life”), which was left blank on 18.9% of questionnaires. On one survey, a respondent wrote “None” in the margin next to question 17 and left the answer blank, another wrote “I don’t have one.” Several respondents chose to write “N/A” next to the work/school items and the relationship items (namely items 7 and 17). Some individuals marked “N/A” next to the alcohol/drug and the relationship questions while still marking answers on the alcohol/drug questions, but leaving the relationship answers blank. This would indicate that respondents understood the directions for answering these questions differently depending on the nature of the item.

Test-Retest Reliability

Hypothesis 2: The test-retest reliability estimate of the OQ-45 with older adults will not be statistically different from test-retest estimates reported in the younger adult age groups.

The test-retest analysis of data collected in this study for the community sample ($r = .91$; $p < .05$; $n = 34$; mean age of 70.12) was consistent with previously published analysis by Lambert et al. (2004) with a student population study ($r = .84$; $n = 157$; mean

age of 23.04). This would indicate that the OQ-45 generally remains stable between first and second administrations with older adults as has been found in previous research with younger adults. However, considering the high correlation estimate over nearly a 3-week interval, it could be argued that the OQ-45 may lack sufficient within-item variance to effectively measure change as a response to therapy. Such a high test-retest correlation could be indicative that the OQ-45 is measuring a construct, such as a highly stable psychological trait, that does not change over time in response to treatment. A defining feature of a trait is stability over time, whereas a state reflects immediate feelings at the point of assessment.

Notably, the OQ-45 has been compared to the State-Trait Anxiety Inventory (STAI) and found to be more highly correlated with “general” feelings of Trait Anxiety (STAI-T; $r = .80$) than with immediate feelings of State Anxiety (STAI-S; $r = .64$; Lambert, Burlingame, Umphress, Hansen, Vermeersch, et al., 1996; Spielberger, 2008), which raises the question of whether the OQ-45 measures a general trait as opposed to an immediate state of psychological distress.

The results of these studies (Lambert et al., 1996; Spielberger, 2008) appear to support the contention that the OQ-45 is measuring a psychological trait that is stable over time rather than a psychological state. This result is understandable considering the nature of the STAI-T as measuring general feelings of anxiety and accompanying feelings also associated with depression (i.e., “I feel inadequate”; Bieling, Antony, & Swinson, 1998). It is also expected, considering that the instructions of the OQ-45 explicitly ask the respondent to answer according to how he or she has been feeling “over the past week,” similar to the instructions on the STAI-T scale of responding to how one

feels “generally” as opposed to the STAI-S scale of responding to how one feels “right at this moment.” This evidence suggesting that the OQ-45 measures a psychological trait raises some concern that OQ-45 scores may be measuring a construct that is too stable to use the instrument to gauge change in distress over time.

Hypothesis 3: As with previous research, scores on the OQ-45 will decrease between first and second administrations of the OQ-45, though this difference will be small and will be less than the Reliable Change Index (RCI) of 14 points.

**Individuals who present with physical health problems
will not demonstrate this decrease.**

In this study, 34 out of 66 community participants completed the follow-up questionnaire. As predicted, scores decreased between administrations by a small margin (average decrease of 1.44 points), though the decrease was not statistically significant. This suggests that the OQ-45 is a stable measure of distress and the test-retest reliability estimate supports this. On average, participant scores on the OQ-45 do not decline substantially over a 3-week period. As noted previously in this manuscript, research with a younger adult population found a statistically significant decrease in scores between first and second administrations. The smaller sample size in this study yields a wider confidence interval and greater standard error for both the test-retest correlation analysis and the means comparison between first and second administrations. Thus the data from this study does not suggest that the instrument is necessarily more stable over time for an older adult group, as the means comparison alone suggests. Notably, the previously published research included a larger sample size and thus was more sensitive to detecting even small changes as statistically significant (Durham et al., 2002). In both this study

and in previously published research, the decrease between first and second administrations did not approach the 14-point change necessary to be considered “clinically significant” by the publishers (Lambert et al., 2004).

Concurrent and Construct Validity

Hypothesis 4: The OQ-45 correlates with measures of: (a) depression, (b) anxiety, and (c) general quality of life. Correlations found in younger populations with similar instruments will fall within the 95% confidence Intervals of correlations found with older adults.

Construct validity of the OQ-45 was examined through correlations with other instruments known to measure similar constructs. The OQ-45 was correlated with measures of depression (GDS), anxiety (GAI), and general quality of life (SF-12). The correlations between the OQ-45 and measures of depression, anxiety, and general mental health, while statistically significant, were smaller in magnitude than those reported in younger adult groups. The 95% confidence intervals of correlation estimates in this study encompassed correlation estimates from previous studies with the OQ-45 and the Beck Depression Inventory (BDI) and the State-Trait Anxiety Inventory- State scale (STAI-S), but other correlation estimates from previous research were outside of the confidence intervals in this study (see Table 8). This would suggest that the OQ-45 is not as highly correlated with measures of depression, anxiety, and general quality of life as has been found in younger adult groups. However, a direct comparison of these correlation estimates must be made with caution because the measures used in this study, while conceptually similar, are not identical to those used in previously published research.

Additionally, the small sample size of this study could have affected the strength of the correlations.

As the contrasting measures used in this study have been evaluated in older adult groups in the published research, and given the correlations found in this study, the OQ-45 can be considered to measure general feelings of psychological distress as reflected through mental health symptoms, which includes self-reported symptoms of depression and anxiety.

Hypothesis 5: Older adults who are receiving treatment for mental health issues in this sample will score significantly higher on the OQ-45 than older adults found in a general community sample.

The OQ-45 appears to differentiate between clinical and community samples of older adults. Clinical participants scored significantly higher than community participants. These results are similar to those found in previous research with younger adults. This would indicate that the OQ-45, when used with an older adult population, may have value in identifying levels of distress among older clients. The correlations between the OQ-45 and measures of depression, anxiety, and general quality of life in combination with the ability to distinguish between clinical and community samples of older adults, provides preliminary evidence of the construct validity of the OQ-45 as a measure of psychological distress.

Despite the similarities in mean scores between this and previous studies, the analysis of this study yielded a different cutoff score (54) for significant distress than found in previous research (64; Lambert et al., 2004). This difference in computed cutoff score resulted in lower Sensitivity (.71) and Specificity (.70) in this study as opposed to

previously (.84 and .83, respectively). This difference could be attributed to differences in standard deviations between these samples and those found in younger adult groups as standard deviations are a critical component of the cutoff score formula. The clinical sample included in this study was characterized by more variability than found in the community sample and also in clinical groups reported in previously published research.

Limitations of This Study

Several study limitations must be considered when interpreting the results of this study including sample size, sampling effects, and design limitations.

Due to limited resources for this study, sample size was a limitation. The community sample consisted of 66 participants, as opposed to a clinical sample of 611 who were actively engaged in mental health treatment at the time of the OQ-45 administration. Of those 66 community participants, only 34 completed the follow-up survey, 5 of whom reported that they were receiving treatment for mental health issues. This small sample size of the community group increased estimates of standard error in the analyses. This statistical issue limits the extent to which these results can be generalized to the broader older adult population as this small sample could differ from the larger population in some manner undetectable by this study. A larger sample of older adults in the community could reduce this likelihood and provide for more robust statistical analyses. Furthermore, limited demographic data of the clinical sample made it difficult to fully characterize the sample. It is possible that the two samples may have differed in some meaningful way other than mental health concerns.

An additional limit to generalizability of these results is potential sampling error. This study employed a convenience sampling strategy. In other words, surveys were

handed out at educational workshops accessible to this researcher. It is not unreasonable to suggest that initial scores may have been affected by participants having very recently engaged in the social events and learning at the workshops. Thus the context in which the surveys were handed out may affect the results of the surveys. For example, only persons who felt that their emotional state was healthy may have volunteered. The community sample consisted of workshop attendees seeking to improve their knowledge about aging issues. Thus they were not representative of the general population of older adults in the community. Furthermore, this study was conducted in a limited geographic area. It is possible that this sample may not be representative of older adults nationwide due to cultural and regional factors.

Finally, this study was limited in scope to the selected aspects of reliability and validity of the OQ-45 scores in the older adult population. Specifically, the concurrent validity estimates from this study may not be trustworthy for making final conclusions and comparisons to similar estimates found in previously published research with a younger adult population. Previous research has evaluated the correlations between the OQ-45 and various measures of depression (BDI, ZSDS), anxiety (ZSAS, STAI-T, STAI-S), and general quality of life (SF-36) with younger adults (Lambert et al., 2004), whereas this study utilized a different set of measures to establish concurrent validity estimates.

For the purpose of this study, instrument selection was balanced between creating a meaningful statistical comparison with previous research (i.e., by using the same instruments as previous research) versus employing the most conceptually applicable instruments (i.e., those developed specifically for older adults). This study could be

viewed as pilot validity data for the OQ-45 with older adults. Practical factors such as ease of administration and participant fatigue were considered in the instrument selection in this study (i.e., the SF-12 was selected instead of the SF-36 to reduce length of the questionnaires). This means that the comparison of the concurrent validity estimates may not be as stable as those that could have been collected using directly equivalent instruments (i.e., the full SF-36, the BDI, etc.). While the results of this study lend some support to the conclusion that the OQ-45 is measuring psychological distress in older adults, it is not possible to conclude that the OQ-45 is performing identically as in younger adult groups. Despite these limitations, this study raises additional questions for further study of the OQ-45 psychometric properties in an older adult sample. These questions are highlighted next.

Recommendations for Future Research

The following three four questions deserve additional study: (1) Does the OQ-45 perform similarly in a larger, more generalizable sample of older adults, and how do these reliability estimates directly compare to those found in younger adults? (2) How does the OQ-45 compare to other concurrent validity measures in older adults? And (3) Is the OQ-45 sensitive to change in psychological distress over time due to clinical interventions with older adults?

While the results of this study lend some support to the use of the OQ-45 with older adults, replication and expansion of this study could strengthen these results or provide useful information regarding the applicability of these results to the larger older adult population. Additional research with a more broad system of recruitment, perhaps through mailing lists, advertising, and recruitment in a broader range of contexts (at

senior apartments, senior centers, advertising through newsletters, random selection in the phonebook, etc.) could yield more generalizable information about the OQ-45 and potentially establish more accurate population estimates for this instrument.

Furthermore, such a study that also concurrently gathers data with a younger adult sample could allow for a direct comparison of OQ-45 scores, changes in scores over time, and item analysis, thus providing useful information about the differences between younger adults and older adults with regards to reported psychological distress.

Future research should also further explore the concurrent validity of the OQ-45 utilizing both measures developed for older adults and those used in general adult populations, which may be helpful to draw more conclusions about the concurrent validity of the OQ-45. For example, one study that administers the OQ-45, BDI, and GDS to the same sample may reveal useful information about how to interpret these instruments accurately with older adults. Such research could also examine the discriminant validity of the OQ-45 with respect to examining the effects of social desirability on OQ-45 scores, as this remains a potential source of variance in self-report instruments. Additional research may also examine the factor structure of these instruments, particularly the OQ-45, as previous research has noted that the factor structure of the OQ-45 may vary with different groups (Lambert et al., 2004). Utilizing qualitative elements (such as focus groups) in research may be helpful in providing feedback regarding instrument construction and assist with identifying problematic items and how to best re-structure or re-phrase items to improve comprehension and applicability of items.

Finally, additional research is also needed to examine changes in OQ-45 scores over time as a response to clinical interventions. While the OQ-45 may be useful as a measure of psychological distress in older adults, as noted in the test-retest analysis, scores on this instrument may reflect a construct that is too stable to be responsive to treatment interventions. Additional research is necessary that examines how participation in treatment may affect reported psychological distress changes over time in a group of older adults. Such a study examining the responsiveness of the OQ-45 to changes due to therapy would allow for a better understanding of the usefulness of the instrument for monitoring outcomes and week-to-week progress in therapy.

Recommendations for Clinical Use

The results of this study would lend some support to the argument that the OQ-45 may be useful as an instrument to measure general feelings of psychological distress in an older adult population. As this study has found the OQ-45 scores to be internally consistent and correlated with other measures of psychological functioning in older adults, the OQ-45 performs similarly with older adults with regards to some estimates of reliability and validity as has been reported in previous research with younger adults. The preliminary evidence would suggest that the OQ-45 also may be useful as a screening tool, though it is not as sensitive or specific as with a younger adult population and a proper cutoff score needs to be established with a larger normative sample before this instrument can be utilized appropriately in this way. In light of the results of this study, some recommendations can be made with regards to how to best utilize the OQ-45 in clinical settings with older adults including (1) revising some of the instructions of the OQ-45 and problematic items, (2) utilizing appropriate cutoff scores based on research

with this population, and (3) appropriately interpreting OQ-45 scores within the clinical context.

Given the high prevalence of missing values on some items, it may be advisable to revise the instructions for the OQ-45 to make it more easily understood, particularly with regards to questions that may be interpreted quite differently by older adults. Creating more specific instructions regarding some of these questions may make the survey more readily understood for older adults who present with many possible answers to the question of relationship status. For example, the instructions could contain directions for answering questions about romantic relationships for those who may be widowed, divorced, or single. Additionally, rephrasing some problematic items could also help improve response rates on certain items. Modifying the work/school questions to include “hobbies” or “daily tasks” as well as “work/school” could increase the response rate for such items, particularly for those who do not work or attend school. The items pertaining to romantic relationships could be modified to mirror the substance abuse items (i.e., “If you are not in a romantic relationship, mark Never”).

The use of appropriate cutoff scores is essential for proper application of the OQ-45 with older adult groups. Utilizing the previously published cutoff score of 64 established with younger adult samples would yield a Sensitivity of .57 and a Specificity of .85. Thus, utilizing the current published cutoff score of 64 would prove problematic if used as a screening tool to determine if someone should be eligible for treatment, as this score in the current study only correctly identified 57% of participants as experiencing clinically significant distress. However, using a cutoff score of 54 (computed using the formula recommended by Lambert et al., 2004) successfully

identified 71% of participants as experiencing clinically significant distress. This underscores the importance of careful consideration when interpreting OQ-45 scores in an older adult population and establishing the appropriate metrics for clinical decision-making.

Primarily, it is recommended that clinicians assess factors beyond the OQ-45 score itself in understanding presenting issues with older adults. While the OQ-45 score may be a useful gauge of general psychological distress, individual items may be particularly useful for engaging in clinical dialogue about the particular concerns with which a client may present. Clinicians may consider asking clients when questions are left blank, which may prompt further discussion into this area of life and promote a better understanding of the experience and feelings of an individual client. For example, if an individual leaves all relationship questions blank, the clinician could query the client further about why the client omitted these questions. If warranted, this could then be a starting point for further dialogue about this topic. As a case in point, while the items pertaining to substance use are problematic from a statistical point of view, they may still have clinical utility as a means of screening for and identifying potential areas of clinical focus on substance use or abuse, particularly if they are all left unanswered by the respondent.

As with all psychological assessment instruments, the information from the OQ-45 is only useful as a clinical tool to supplement the therapy process. The results of this study suggest that when using the OQ-45 with older adults, the recommendations of the authors of the OQ-45 remain useful. Questionnaires should not be exclusively relied on in clinical decision making, but should be part of a more comprehensive assessment

inclusive of a direct one-on-one clinical interview with a qualified mental health professional, ideally employing behavioral data as well as information from collateral sources such as patient family, previous psychotherapy, and medical history.

REFERENCES

- Ahava, G.W., Iannone, C., Grebstein, L., & Schirling, J. (1998). Is the Beck Depression Inventory reliable over time? An evaluation of multiple test-retest reliability in a nonclinical college student sample. *Journal of Personality Assessment, 70*(2), 222-231).
- Alexopoulos, G.S. (2005). Depression in the elderly. *The Lancet, 365*, 1961-1970.
- American Psychological Association. (2004). Report of the association: Guidelines for psychological practice with older adults. *American Psychologist, 59*(4), 236-260.
- Anderson, T., Ogles, B.M., Patterson, C.L., Lambert, M.J., & Vermeersch, D.A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755-768.
- Beckstead, D.J., Hatch, A.L., Lambert, M.J., Eggett, D.L., Goates, M.K., & Vermeersch, D.A. (2003). Clinical significance of the Outcome Questionnaire (OQ45.2). *The Behavior Analyst Today, 4*(1), 79-90.
- Carstensen, L.L., Isaacowitz, D.M., & Charles, S.T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54*(3), 165-181.
- Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.
- Doerfler, L.A., Addis, M.E., & Moran, P.W. (2002). Evaluating mental health outcomes in an inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32. *The Journal of Behavioral Health Services & Research, 29*(4), 394-403.
- Durham, C.J., McGrath, L.D., Burlingame, G.M., Schallje, G.B., Lambert, M.J., & Davies, D.R. (2002). The effects of repeated administrations on self-report and parent-report scales. *Journal of Psychoeducational Assessment, 20*, 240-257.
- Edelstein, B., Kalish, K.D., Drozdick, L.W., & McKee, D.R. (2000). Assessment of depression and bereavement in older adults. In P.A. Lichtenberg (Ed.) *Handbook of assessment in clinical gerontology*. (pp. 3-43). New York: John Wiley & Sons, Inc.
- Eisen, S.V, Dill, D.L., & Grob, M.C. (1994). Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. *Hospital and Community Psychiatry, 45*(3), 242-247.

- Finch, A.E., Lambert, M.J., & Schaalje, B.G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy*, 8, 231-242.
- Gallo, J.J. & Bogner, H.R. (2006). The context of geriatric care. In J.J. Gallo, H.R. Bogner, T. Fulmer, & G.J. Paveza (Eds.). *Handbook of geriatric assessment (4th ed.)*. (pp. 3-14). Sudbury, MA: Jones and Bartlett Publishers, Inc.
- Gurland, B.J., Wilder, D.E., & Berkman, C. (1988). Depression and disability in the elderly: Reciprocal relations and changes with age. *International Journal of Geriatric Psychiatry*, 3, 163-179.
- Hayes, W.L. (1994). *Statistics (5th ed)*. Fort Worth, TX: Harcourt Brace College Publishers.
- Hill, R.D. (2005). *Positive aging: A guide for mental health professionals and consumers*. New York: WW Norton.
- Hill, R.D. (2011). A positive aging framework for guiding geropsychology interventions. *Behavior Therapy*, 42, 66-77.
- Hill, R.D., Thorn, B.L., & Packard, T. (2000). Counseling older adults: Theoretical and empirical issues in prevention and intervention. In S.D. Brown & R.W. Lent (Eds.), *Handbook of counseling psychology (3rd ed.)*. (pp. 499-531). New York: John Wiley & Sons, Inc.
- de Jong, K., Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, P., & Heiser, W.J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy*, 14, 288-301.
- Jurado, J. (2007). Towards an understanding of the psychometric properties of the Spanish version of the OQ-45. Unpublished doctoral dissertation, Brigham Young University, Provo, UT. Retrieved November 4, 2009 from ProQuest Digital Dissertations.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook (3rd ed)*. Upper Saddle River, NJ: Prentice Hall.
- LaBuda, J., & Lichtenberg, P. (1999). The role of cognition, depression, and awareness of deficit in predicting geriatric rehabilitation patients' IADL performance. *The Clinical Neuropsychologist*, 13(3), 258-267.
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17(1), 1-14.

- Lambert, M.J., Burlingame, G.M., Umphress, V., Hansen, N.B., Vermeersch, D.A., Clouse, G.C., & Yanchar, S.C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3(4), 249-258.
- Lambert, M.J., Morton, J.J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R.C., et al. (2004). *Administration and scoring manual for the Outcome Questionnaire 45*. Salt Lake City, UT: OQ Measures L.L.C.
- Lo Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, M.J. (2008). The factorial structure of the Outcome Questionnaire-45: A study with an Italian sample. *Clinical Psychology and Psychotherapy*, 15, 418-423.
- Lueck, W.C. (2004). Differential treatment response as a function of diagnostic classification. Unpublished doctoral dissertation, Brigham Young University, Provo, UT. Retrieved November 4, 2009 from ProQuest Digital Dissertations.
- Minami, T., Tierney, S.C., Benbrook, A.R., Davies, D.R., Bettman, J.E., McAward, S.M. ... Wampold, B.E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology*, 56(2), 309-320.
- Mueller, R.M., Lambert, M.J., & Burlingame, G.M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, 70(2), 248-262.
- Pachana, N.A., Byrne, G.J., Siddle, H., Koloski, N., Harley, E., & Arnold, E. (2007). Development and validation of the Geriatric Anxiety Inventory. *International Psychogeriatrics*, 19(1), 103-114.
- Pienta, A.M., Hayward, M.D., & Jenkins, K.R. (2000). Health consequences of marriage for the retirement years. *Journal of Family Issues*, 21(5), 559-586.
- Rogers, A. & Delewski, C. (2004). Elders with serious and persistent mental illness (SPMI). *Journal of Gerontological Social Work*, 43(1), 19-36.
- Roseborough, D.J., Luptak, M., McLeod, J., & Bradshaw, W. (2013). Effectiveness of psychodynamic psychotherapy with older adults: A longitudinal study. *Clinical Gerontologist*, 36(1), 1-16.
- Scogin, F. (2007). Introduction to the special section on evidence-based psychological treatments for older adults. *Psychology and Aging*, 22(1), 1-3.
- Spielberger, C.D. (2008). The State-Trait Anxiety Inventory for Adults: The most widely used self-report measure of anxiety. Retrieved February 23, 2010 from STAI website: <http://www.mindgarden.com/products/staisad.htm>.

- Stephens, C., Alpass, F., Towers, A., & Stevenson, B. (2011). The effects of types of social networks, perceived social support, and loneliness on the health of older people: Accounting for the social context. *Journal of Aging and Health, 23*(6), 887-911.
- Vermeersch, D.A., Lambert, M.J., & Burlingame, G.M. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74*(2), 242-261.
- Vogeli, C., Shields, A.E., Lee, T.A., Gibson, T.B., Marder, W.D., Weiss, K.B., & Blumenthal, D. (2007). Multiple chronic conditions: Prevalence, health consequences, and implications for quality, care management, and costs. *Journal of General Internal Medicine, 22*(3), 391-395.
- Whipple, J.L., Lambert, M.J., Vermeersch, D.A., Smart, D.W., Nielsen, S.L., & Hawkins, E.J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problems-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59-68.
- Wrosch, C., Dunne, E., Scheier, M.F., & Schulz, R. (2006). Self-regulation of common age-related challenges: Benefits for older adults' psychological and physical health. *Journal of Behavioral Medicine, 29*(3), 299-306.
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M.B., Leirer, V.O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*, 37-49.
- Yochim, B.P., Lequerica, A., MacNeill, S.E., Lichtenberg, P.A. (2008). Cognitive initiation and depression as predictors of future instrumental activities of daily living among older medical rehabilitation patients. *Journal of Clinical and Experimental Neuropsychology, 30*(2), 236-244.
- Yon, A., & Scogin, F. (2007). Procedures for identifying evidence-based psychological treatments for older adults. *Psychology and Aging, 22*(1), 4-7.
- Ware, J.E., Kosinski, M.M., & Keller, S.D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.