



Published in final edited form as:

Qual Life Res. 2017 November ; 26(11): 2987–2997. doi:10.1007/s11136-017-1631-4.

Viability of the World Health Organization Quality of Life Measure to Assess Changes in Quality of Life following Treatment for Alcohol Use Disorder

Megan Kirouac, Elizabeth R. Stein, Matthew R. Pearson, and Katie Witkiewitz

Department of Psychology, Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico

Abstract

Purpose—Quality of life is an outcome often examined in treatment research contexts such as biomedical trials, but has been studied less often in alcohol use disorder (AUD) treatment. The importance of considering QoL in substance use treatment research has recently been voiced, and measures of QoL have been administered in large AUD treatment trials. Yet, the viability of popular QoL measures has never been evaluated in AUD treatment samples. Accordingly, the present manuscript describes a psychometric examination of and prospective changes in the World Health Organization Quality of Life measure (WHOQOL-BREF) in a large sample ($n=1383$) of patients with AUD recruited for the COMBINE Study.

Methods—Specifically, we examined the construct validity (via confirmatory factor analyses), measurement invariance across time, internal consistency reliability, convergent validity, and effect sizes of post-treatment changes in the WHOQOL-BREF.

Results—Confirmatory factor analyses of the WHOQOL-BREF provided acceptable fit to the current data and this model was invariant across time. Internal consistency reliability was excellent ($\alpha > .9$) for the full WHOQOL-BREF for each timepoint, the WHOQOL-BREF had good convergent validity, and medium effect size improvements were found in the full COMBINE sample across time.

Conclusions—These findings suggest the WHOQOL-BREF is an appropriate measure to use in samples with AUD, that the WHOQOL-BREF scores may be examined over time (e.g., from pre- to post-treatment), and the WHOQOL-BREF may be used to assess improvements in quality of life in AUD research.

Keywords

quality of life; measurement invariance; confirmatory factor analysis; effect size; alcohol use disorder; alcohol dependence

Corresponding Author: Megan Kirouac, Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico, 2650 Yale Blvd SE, MSC 11-6280, Albuquerque, NM 87106. mkirouac@unm.edu.

No authors have any conflicts of interest.

Introduction

Prior research on the treatment of alcohol use disorder (AUD) has predominantly evaluated treatment efficacy via consumption-based outcomes (i.e., endpoints), such as percent days abstinent or percent subjects with no heavy drinking days (e.g., Falk et al., 2010). The development and maintenance of AUD are inherently tied to alcohol consumption itself, yet the sole reliance on consumption-based outcomes as the indicators of treatment efficacy has numerous limitations. Perhaps most notable of these limitations is the fact that consumption is not the only clinically relevant outcome. For example, consumption is not part of the diagnostic criteria for AUD for World Health Organization (WHO) International Classification of Diseases (WHO, 2015) or Diagnostic and Statistical Manual for Mental Disorders, 5th edition (American Psychiatric Association, 2013). Second, consumption is not all that matters to clients, clients' loved ones, or clinicians (Kaskutas et al., 2014; Neale et al., 2014).

Kaskutas and colleagues (2014) surveyed over 9,300 individuals and found that non-consumption outcomes are critical for evaluating client treatment benefit, rather than examining reductions in consumption alone. Similarly, treatment providers in the UK reported that treatment benefit should be based on a variety of outcomes, including psychological well-being, physical health, and social functioning (Neale et al., 2014). These non-consumption variables have been previously conceptualized as the components of overall quality of life (QoL; e.g., Skevington et al., 2004).

Not surprisingly, QoL has been proposed as a non-consumption outcome to be evaluated systematically in AUD trials (Donovan et al., 2005). Evaluating treatment benefit based on improvements in QoL would be an important shift in AUD treatment research not only because it is more clinically meaningful, but it would also apply to a variety of treatment goals, including abstinence and moderation goals (Donovan et al., 2005; Marlatt & Witkiewitz, 2010). One criticism of relying on consumption outcomes as the only acceptable outcomes for establishing treatment efficacy is that such an outcome (e.g., abstinence) may not be consistent with individual client goals (e.g., Donovan et al., 2005; Donovan et al., 2012). Examining QoL as an outcome within the AUD treatment framework allows clinicians and researchers to broaden their definition of client improvement to include non-consumption outcomes, which could be particularly useful when working with individuals for whom abstinence treatment goals are not desirable.

Evaluating treatment benefit based on QoL rather than consumption, is also consistent with studies of other psychological disorders and medical conditions. For example, biomedical research has shifted to considering overall QoL as a primary outcome variable (Donovan et al., 2005). Unsurprisingly, there is considerable research utilizing QoL measures across adult populations among individuals with and without physical or mental illnesses (e.g., Chen, Wu, & Yao, 2006; Ohaeri et al, 2007; Trompenaars et al., 2005; Yao & Wu, 2005; Zubaran & Foresti, 2009). There have also been a few studies of QoL among individuals with AUD and other addictions (e.g., François et al., 2015; Luquiens et al., 2012; Tracy et al., 2012).

Despite the recommendations to examine QoL and some previous research examining QoL as a measure of treatment efficacy in AUD samples, QoL has yet to be examined systematically in AUD treatment research. The lack of systematic evaluation of QoL is problematic because the lack of consistency in the QoL measures used across trials makes cross-study comparisons difficult, and, in other trials, QoL may be overlooked completely (Luquiens et al., 2012). One potential reason for the delay in shifting toward examining QoL in AUD treatment research may be that no “gold standard” measure of QoL has been identified for use in AUD patients. Nonetheless, several self-report and observer report measures of QoL have been developed and examined in multiple populations (The EuroQoL Group, 1990; International Resource Center for Health Care Assessment, 1992; WHOQOL Group, 1998), but are rarely studied or reported in AUD samples (Donovan et al., 2005). Therefore, we do not know if extant QoL measures are appropriate to use in this population. Moreover, no studies have examined measurement invariance of any QoL measures in an AUD sample to evaluate the appropriateness of comparing scores across time (e.g., pre to post treatment) of QoL measures to test treatment effects. Finally, identifying a possible gold-standard QoL measure in AUD populations might allow future researchers to administer that measure, thus facilitating treatment efficacy comparisons between studies via integrative data analysis (e.g., Curran et al., 2014). Accordingly, the present study aimed to address this important gap by examining the viability of one of the most widely-used QoL measures (the World Health Organization’s Quality of Life-Brief version) for use in AUD samples. Specifically, we used data collected in the COMBINE Study to explore if the WHOQOL-BREF might be viable for evaluating QoL in AUD samples by examining several aspects of the WHOQOL-BREF: construct validity (via confirmatory factor analysis), measurement invariance, internal consistency reliability, convergent validity, and ability to detect changes in QoL over time (via Cohen’s *d* effect sizes).

Method

Participants and Procedures

The present analyses used data from the COMBINE Study ($N=1383$; COMBINE Study Group, 2003), a multisite randomized clinical trial for individuals who met criteria for alcohol dependence (DSM-IV-TR; American Psychiatric Association, 2000). Participants were 69.1% male; 76.8% non-Hispanic, White; 11.2% Hispanic; 7.9% Black; 1.3% American Indian; and 2.8% were other or mixed race; and participants’ average age was 44.4 years ($SD=10.2$). Participants were randomized to one of nine treatment cells using a 2 (active naltrexone versus placebo naltrexone) \times 2 (active acamprosate versus placebo acamprosate) \times 2 (Medication Management (MM) versus Combined Behavioral Intervention (CBI)) + 1 (CBI only with no pills) design. All participants received treatment for 16 weeks and completed follow-up assessments for up to 1 year following treatment. More information on study design and results have been published previously (Anton et al., 2006; Donovan et al., 2008).

Quality of Life

The World Health Organization (WHO) developed a measure of quality of life (WHOQOL; WHOQOL Group, 1998). The WHOQOL is considered a generic, multidimensional scale

because it was not designed to be used with any particular population and it covers a broad spectrum of dimensions of QoL (e.g., physical, psychological, and social health; Zubaran & Foresti, 2009). To reduce respondent burden, 26 items were extracted from the 100-item WHOQOL in the formation of the WHOQOL-BREF (Lucas-Carrasco, Laidlaw, & Power, 2011; Skevington et al., 2004).

The WHOQOL-BREF was developed using a unique, cross-cultural methodology, and its psychometric properties have been studied in many distinct cultures (e.g. Dutch, Trompenaars et al., 2005; Sudanese, Ohaeri et al, 2007; Taiwanese, Yao & Wu, 2005). Several studies have examined the psychometric properties of the WHOQOL-BREF (e.g., Jaracz et al., 2006, Yao & Wu, 2002). For example, Skevington and colleagues (2004) examined the psychometric properties of the WHOQOL-BREF using data collected from a survey of adults in 23 countries and in both clinical and non-clinical (i.e., healthy) populations. The WHOQOL-BREF was determined to have good-to-excellent internal consistency and was found to have a higher-order factor structure consisting of four lower order factors (physical health, psychological health, social relationships, and environment) and one higher-order factor for overall QoL (Skevington et al., 2004).

In the COMBINE Study, a 25-item assessment of various aspects of QoL based on the WHOQOL-BREF (WHOQOL Group, 1998) was administered. One item that assessed the presence of negative feelings was unintentionally excluded from the original measure in the COMBINE assessment battery. Consistent with previous factor analyses of the WHOQOL-BREF, two items assessing overall QoL were excluded from the present analyses (see Table 1 for a brief description of the items examined in the present study; Skevington et al., 2004). Response options in the COMBINE study were ordered categorical, ranging from 1 (“not at all”) to 5 (“an extreme amount”). For the present analyses, items were recoded using the original categorical options (WHOQOL Group, 1998) so that higher scores on each individual item consistently indicated better QoL. The WHOQOL-BREF was administered at baseline, 10 weeks following treatment (week 26) and 36 weeks following treatment (week 52).

Data Preparation and Analyses

Figure 1 presents a visual depiction of data preparation and analysis plans. Preliminary data screening for the WHOQOL-BREF items in the COMBINE Study indicated all item pairwise correlations were less than 0.8. Visual inspection of histograms of data suggested data were mostly distributed within normal range and data existed within each categorical response option for the whole sample. No data transformations of the WHOQOL-BREF were used for the present analyses and all analyses considered the categorical nature of the items.

Construct Validity Analyses—Since multiple factor structures are published, confirmatory factor analyses were used to test these previously published factor structures in COMBINE. To avoid getting an adequately fitting model due to chance alone, we split the COMBINE study sample with available data on the WHOQOL ($n=1351$) using SPSS version 23 (IBM Corp, 2015) into two independent subsamples ($n=672$; $n=679$) by randomly

selecting approximately 50% of cases for a development sample (Sample 1, $n=672$) and approximately 50% of cases for a replication sample (Sample 2, $n=679$). We then used Sample 1 to specify and respecify the previously published factor models. Once we selected a final factor model using Sample 1 (i.e., the development sample), we then replicated the final model in Sample 2. All factor analyses were performed using Mplus version 7.3 (Muthén & Muthén, 2012) with the mean- and –variance adjusted weighted least squares (WLSMV) estimator since response options were categorical. Moreover, because there were significant differences in participant demographics by treatment site in the COMBINE study, all analyses used the WLSMV estimator to adjust the standard errors for clustering within treatment sites (Muthén & Muthén, 2012).

CFA was used to evaluate the factor structure of the WHOQOL-BREF to evaluate construct validity. The CFA model testing was based on the widely cited factor structure described by Skevington and colleagues (2004), shown in Figure 2. We also tested other widely cited published factor structures to compare model fit to alternatives (Jaracz et al., 2006; Trompenaars et al., 2005; Yao & Wu, 2002). *A priori* criteria for acceptable model fit were defined by comparative fit index (CFI) $\geq .9$ and root mean square error of approximation (RMSEA) $\leq .08$, which is consistent with recommendations for adequate, but not strong model fit (Hu & Bentler, 1999). Standardized factor loadings $>.30$ were considered adequate, given the primary goal of the study was to examine the fit and measurement properties of an established factor structure in an AUD sample, rather than propose a new factor structure for the WHOQOL-BREF.

Measurement Invariance Analyses—To examine measurement invariance of the WHOQOL-BREF across time (baseline and week 26, and again between weeks 26 and 52), we employed the model-based technique described by Chen and colleagues (2005) for testing invariance of higher-order factor structures, where overall QoL was the higher order factor and physical health, psychological health, social relationships, and environment were the lower order factors. Briefly, configural invariance was tested first across time by freely estimating all parameters at each of the timepoints (baseline; week 26; and week 52 - Model 1). Second, metric invariance (i.e., invariance of the factor loadings) was measured across time using two different models. For Model 2, factor loadings for the lower-order factors were constrained to be equivalent across time. For Model 3, factor loadings were constrained at the higher-order factor to be equal across time. Models 1–3 were all nested and invariance was supported if the change in CFI from Model 1 to Model 2 and Model 2 to Model 3 was less than .01 as recommended by Cheung and Rensvold (1999). We relied on the change in CFI, rather than χ^2 difference testing, to assess measurement invariance testing because the χ^2 difference test is often unfairly biased by large sample sizes (Cheung and Rensvold, 2002; Widaman et al., 2010).

Due to the categorical response options for the WHOQOL-BREF items, the tests of scalar invariance (i.e., invariance of the item thresholds and factor intercepts) required additional model constraints for model identification. Specifically, item residual variances were constrained to 1 and factor means were constrained to 0 in the first timepoint (e.g., baseline) for Model 4. Then, factor means were constrained to 0 for both timepoints (e.g., baseline and week 26) in Model 5. Consequently, Model 4 built upon Model 3 by adding the

additional constraint of equivalent item thresholds for the lower-order factors and Model 5 added the constraint of equivalent factor intercepts for the higher-order factor. However, since Models 4 and 5 deviated from the nested model structure used in Models 1 through 3, determining time invariance was based on a priori cutoffs for acceptable fit indices (CFI .9 and RMSEA .08) and the change in CFI from Model 4 to Model 5. Due to problems with model identification and to recommendations in the literature (Cheung & Rensvold, 2002), residual invariance (i.e., “strict invariance”) was not tested in the present analyses. The full sample size was used for all invariance testing.

Convergent Validity Analyses—Convergent validity of the WHOQOL-BREF was tested via bivariate correlations between the total WHOQOL-BREF score and subscale scores with other, related measures. These scores on the baseline WHOQOL-BREF were examined in relation to baseline scores on the Alcohol Dependence Scale (ADS; Skinner & Allen, 1982), the Brief Symptom Inventory (BSI; Derogatis, 1993), and the Drinker Inventory of Consequences (DrInC; Miller et al., 1995). The ADS is a 25-item assessment of symptoms of alcohol dependence; we used the total score in the present study. Internal consistency of baseline ADS was $\alpha=0.849$. The BSI is a 53-item measure of general psychiatric symptoms; internal consistency of baseline BSI was $\alpha=0.965$. We used both the total BSI score (representing global psychological problem severity) and scores on each of the nine subscales. The DrInC is a 45-item measure of alcohol-related consequences; internal consistency of baseline DrInC was $\alpha=0.937$. The DrInC assesses frequency of experience for each consequence in the assessment window (responses are 0=“never” to 3=“daily or almost daily”). The present study examined total DrInC score (excluding the control-scale items) as well as scores on each of the 5 subscales. Since higher scores on the WHOQOL-BREF indicated better QoL and higher scores on the ADS, BSI, and DrInC (including their subscales) indicate poorer functioning (i.e., higher problem severity), all bivariate correlations were hypothesized to be negative.

Internal Consistency Reliability and Effect Size Analyses—Internal consistency reliability was analyzed using Cronbach’s alpha via version 23 of SPSS (IBM Corp, 2015). *A priori* cutoffs for internal consistency reliability were: >0.9 as “excellent,” >0.8 as “good,” and >0.7 as “acceptable” (Gliem & Gliem, 2003). Effect sizes were calculated via Cohen’s *d* in SPSS version 23 (IBM Corp, 2015). *A priori* cutoffs for effect sizes were: >0.6 as “large,” 0.3 to 0.6 as “medium,” and <0.3 as “small.” We evaluated effect sizes for the changes in average scores for each subscale (physical health, psychological health, social relationships, and environment) and for total WHOQOL-BREF summary scores in the full COMBINE sample and in the two sub-samples that had greatest changes in abstinence rates: naltrexone versus placebo and CBI versus MM (Anton et al., 2006).

Results

Model results from the confirmatory factor analyses suggested that none of the tested models fit exceptionally better than any other and one model was non-positive definite (the model based on Yao & Wu, 2002). Table 2 presents the fit indices for each of the tested models. Since the model specified by Skevington et al. (2004), shown in Figure 2, makes conceptual sense with factors comprising QoL and since that publication was the most

widely cited of the tested models, we choose to proceed with the model specified by Skevington et al. (2004) as one that may be most useful. We tested a factor model consisting of physical health, psychological health, social relationships, and environment as lower order factors with one higher-order overall QoL factor that contained the items described by Skevington and colleagues (2004). CFA of the 23 item COMBINE study WHOQOL-BREF at the baseline assessment indicated acceptable model fit of both split samples (development sample: $n=672$, $RMSEA=0.054$ (90% CI: 0.050, 0.059), $CFI=0.940$; replication sample: $n=679$, $RMSEA=0.050$ (90% CI: 0.045, 0.055), $CFI=0.942$). Table 1 presents the items by factor and their respective factor loadings. Further, each first-order factor had high standardized factor loadings on the higher-order QoL factor (see Figure 2).

Next, we tested for factorial invariance across timepoints using the full sample to maximize data availability. The fit indices and change in model fit across Models 1–3 and Models 4–5 are provided in Table 3. As shown in the table, all five models provided an acceptable fit to the data based on the RMSEA and CFI, with acceptable fit of Model 1 at all timepoints providing support for configural invariance. Moreover, nested model comparisons results supported metric invariance of the lower order factor loadings (Model 1 vs. Model 2; baseline to week-26: $CFI=.005$; week-26 to 52: $CFI=.007$), metric invariance of the higher order factor loadings (Model 2 vs. Model 3; baseline to week-26: $CFI=.001$; week-26 to 52: $CFI=.001$), and scalar invariance of the lower order factor means (Model 4 vs. Model 5; baseline to week-26: $CFI=.007$; week-26 to 52: $CFI=.001$). Accordingly, we concluded that the WHOQOL-BREF is invariant across time from baseline to 10 weeks following treatment (week-26) and from 10 weeks following treatment to 36 weeks following treatment (week-52).

The internal consistency reliability of the baseline, week-26, and week-52 data for the WHOQOL-BREF was acceptable (Cronbach $\alpha > 0.70$). For the full WHOQOL-BREF measure, internal consistency reliability at each timepoint was excellent (Cronbach $\alpha > 0.90$). Table 4 presents Cronbach α s for each subscale and total scale for each of the three timepoints. Further, the WHOQOL-BREF total summary score had excellent convergent validity per significant bivariate correlations with all tested measures ($p < .001$; see Table 5). Convergent validity of the total score and the physical health factor subscales of the WHOQOL-BREF was demonstrated by significant correlations with other indices of psychological functioning (i.e., ADS, BSI total and subscale scores, DrInC total and subscale scores). The remaining subscales demonstrated very weak (mostly non-significant) associations with other indices of psychological functioning.

Cohen's d effect sizes comparing subscale and total WHOQOL-BREF summary scores from baseline to the week-26 and week-52 timepoints within the full sample were in the medium range ($d > 0.30$; see Table 6). Accordingly, the WHOQOL-BREF appears able to detect changes in QoL following treatment. Effect sizes comparing subscale and total summary scores of the WHOQOL-BREF at week-26 and week-52 timepoints between treatment conditions are presented in Table 7. All effect sizes within the two treatment comparison sub-groups examined in the present study were in the small range ($d < 0.30$).

Discussion

The present study provides empirical support for the viability of the WHOQOL-BREF in a sample of individuals with AUD. The construct validity was supported in that the widely cited factor structure of the WHOQOL-BREF (Skevington et al., 2004) fit data from COMBINE. Importantly, the present study established measurement invariance of the WHOQOL-BREF across multiple timepoints. Measurement invariance was established for the higher-order factor structure through “strong invariance.” Substantively, these findings indicate that factor scores including the higher-order QoL factor scores may be compared across time (pre- and post-treatment, post-treatment and follow-up) among individuals with AUD. Moreover, internal consistency reliability and convergent validity of the full WHOQOL-BREF were excellent and medium effect sizes were found between baseline and later timepoints within the full sample of COMBINE. These findings contribute important information to the field by identifying the WHOQOL-BREF as a viable measure for evaluating improvements in QoL following treatment for AUD, which is an important indicator of improvement from both client and clinician perspectives (e.g., Kaskutas et al., 2014; Neale et al., 2014). Importantly, the WHOQOL-BREF appears to be a useful measure for demonstrating improvements in QoL following treatment; however, the between treatment group effect sizes at 10-weeks and 9-months post-treatment were quite small and suggest that the WHOQOL-BREF was unable to detect differences between active treatment conditions in the COMBINE study. It is also critical to acknowledge that a lack of difference between groups in QoL measures does not inherently mean QoL measures, and other non-consumption outcomes are necessarily insensitive to detect treatment effects. It may be that QoL is a less proximal outcome than alcohol consumption, that the treatments did not effectively target QoL improvements, or a number of alternative explanations for why we did not find larger effect sizes between active treatment conditions. Future work should continue to examine other QoL and other non-consumption measures to determine whether other measures are more useful as indicators of treatment efficacy in AUD treatment studies.

Limitations and Future Directions

Although the study provides evidence for the utility of the WHOQOL-BREF in AUD treatment research, the present analyses are not without limitations. First, the models tested to establish measurement invariance were not all perfectly nested. Accordingly, fit indices of Models 1–3 cannot be directly compared to those of Models 4–5. Nonetheless, fit indices of Models 1–5 all provided adequate fit per *a priori* fit indices and model fit generally improved as additional constraints were added. Together, these findings support our conclusion of measurement invariance of the WHOQOL-BREF across time in the COMBINE Study.

A second limitation to the present study is that we only evaluated the measurement invariance and psychometric properties for the most widely-cited factor structure of this QoL scale (Skevington et al., 2004). Alternative factor structures fit similarly in the present study and may have also demonstrated measurement invariance and similar psychometric properties. However, the majority of the alternative factor structures are comprised of a four-factor solution that is largely similar to that published by Skevington and colleagues (2004). Further, the factor solution published by Skevington and colleagues (2004) was the most-

widely cited article examining the WHOQOL factor structure at the time of manuscript preparation for the present study. Accordingly, although there are always alternative factor structures that may provide similar or even better model fit, we chose to test the factor structure that has proven most useful in the literature.

Another limitation to the present study is that COMBINE omitted one item of the WHOQOL-BREF that assesses negative affect. Although negative affect may be potentially important to examine in AUD populations, it is important to know that the psychological health factor is robust in the face of this clerical error. Further, the WHOQOL-BREF was highly correlated with other measures of psychological well-being, which may suggest that the WHOQOL-BREF in COMBINE still assesses psychological problems such as those that may be related to negative affect.

It is also important to evaluate other psychometric characteristics of the WHOQOL-BREF in these samples. Specifically, future research should test the sensitivity and specificity of the WHOQOL-BREF for detecting clinically meaningful changes. For example, receiver operating characteristic curves (Hanley & McNeil, 1982) may be used to examine how suitable the WHOQOL-BREF may be as a predictor of other outcomes (e.g., consumption, alcohol-related consequences). Such research would be consistent with recent recommendations to consider a variety of treatment outcomes in evaluating AUD treatment efficacy (e.g., Donovan et al., 2012; Tiffany et al., 2012). Consequently, future research would be able to evaluate treatment efficacy based on more clinically meaningful outcomes than abstinence or other consumption outcomes alone.

Conclusions

Researchers have called for increased use of clinically meaningful outcome variables in evaluating treatment of addictive behaviors beyond abstinence alone (e.g., Midanik, Greenfield, & Bond, 2007; Moos & Finney, 1983; Witkiewitz, 2013). Quality of life (QoL), comprised of physical health, psychological health, social relationships, and environmental factors may be a particularly appropriate non-consumption variable for AUD treatment researchers to use because it assesses various aspects of one's life that have been highlighted as meaningful to clients and their loved ones in addition to treatment providers (Kaskutas et al., 2014; Neale et al., 2014). The findings of the present study suggest the WHOQOL-BREF may be a psychometrically viable measure for AUD treatment researchers to use systematically to compare baseline and post-treatment changes in QoL. Future AUD treatment researchers can examine pre- and post-treatment changes on the WHOQOL-BREF as a concise way to evaluate treatment benefit beyond alcohol use or abstinence alone.

Acknowledgments

This research was supported by a grant (R01-AA022328; PI: Witkiewitz) from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). MK is supported by a fellowship grant (F31-AA024959) from the NIAAA. MRP is supported by a career development grant (K01-AA023233) from the NIAAA.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000. text rev
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM. ... COMBINE Study Research Group. Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence: The COMBINE Study: A Randomized Controlled Trial. *The Journal of the American Medical Association*. 2006; 295(17):2003–2017. [PubMed: 16670409]
- Barros de Silva Lima AF, Fleck M, Pechansky F, de Boni R, Sukop P. Psychometric properties of the World Health Organization Quality of Life instrument (WHOQoL-BRIEF) in alcoholic males: A pilot study. *Quality of Life Research*. 2005; 14:473–478. [PubMed: 15892436]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107(2):238–46. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2320703>. [PubMed: 2320703]
- Bureau of Labor Statistics. U.S. Department of Labor, Occupational Outlook Handbook, 2014–15 Edition. 2014. Retrieved from <http://www.bls.gov/ooh/life-physical-and-social-science/>
- Chen FF, Sousa KH, West SG. Testing measurement invariance of second-order factor models. *Structural Equation Modeling*. 2005; 12(3):471–492.
- Chen KH, Wu CH, Yao G. Applicability of the WHOQOL-BREF on early adolescence. *Social Indicators Research*. 2006; 79:215–234. DOI: 10.1007/s11205-005-0211-0
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 1999; 9(3):233–255.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
- COMBINE Study Group. Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: Rationale and methods. *Alcoholism: Clinical and Experimental Research*. 2003; 27(7):1107–22.
- Derogatis, LR. BSI. Brief Symptom Inventory. Administration, Scoring, and Procedures Manual. 3. National Computer Systems, Inc; Minneapolis, MN: 1993.
- Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, ... Wells EA. Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials. *Addiction*. 2012; 107(4):694–708. [PubMed: 21781202]
- Donovan D, Mattson ME, Cisler RA, Longabaugh R, Zweben A. Quality of life as an outcome measure in alcoholism treatment research. *Journal of Studies on Alcohol*. 2005; S15:119–139.
- The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990; 16:199–208. [PubMed: 10109801]
- Falk D, Wan XQ, Liu L, Fertig J, Mattson M, Ryan M, ... Litten RZ. Percentage of subjects with no heavy drinking days: Evaluation as an efficacy endpoint for alcohol clinical trials. *Alcoholism: Clinical and Experimental Research*. 2010; 34(12):2022–2034. DOI: 10.1111/j.1530-0277.2010.01290.x
- François C, Rahhali N, Chalem Y, Sørensen P, Luquiens A, Aubin HJ. The Effects of as-Needed Nalmefene on Patient-Reported Outcomes and Quality of Life in Relation to a Reduction in Alcohol Consumption in Alcohol-Dependent Patients. *PLoS One*. 2015; 10(8)doi: 10.1371/journal.pone.0129289
- Gliem, JA., Gliem, RR. Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for Likert-type scales. *Midwest Research to Practice Conference in Adult, Continuing, and Community Education*; 2003. p. 82-88. <http://www.ssnpstudents.com/wp/wp-content/uploads/2015/02/Gliem-Gliem.pdf>
- Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
- Huang IC, Wu AW, Frangakis C. Do the SF-26 and WHOQOL-BREF measure the same constructs? Evidence from the Taiwan population. *Quality of Life Research*. 2006; 15:15–24. DOI: 10.1007/s11136-005-8486-9 [PubMed: 16411027]

- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- IBM Corp. IBM SPSS Statistics for Windows. Version 23.0. Armonk, NY: IBM Corp; 2015.
- International Resource Center for Health Care Assessment. *How to Score the SF-36 Short-Form Health Survey*. Boston, MA: The Health Institute; 1992.
- Jaracz K, Kalfoss M, Górna K, Bczyk G. Quality of life in Polish respondents: Psychometric properties of the Polish WHOQOL-BREF. *Scandinavian Journal of Caring Science*. 2006; 20:251–260.
- Kline, RB. *Principles and Practice of Structural Equation Modeling*. 3. New York, NY: The Guilford Press; 2011.
- Luquiens A, Reynaud M, Falissard B, Aubin HJ. Quality of life among alcohol-dependent patients: how satisfactory are the available instruments? A systematic review. *Drug and Alcohol Dependence*. 2012; 125(3):192–202. DOI: 10.1016/j.drugalcdep.2012.08.012 [PubMed: 22939877]
- Marlatt GA, Witkiewitz KA. Harm reduction approaches to alcohol use: Health promotion, prevention, and treatment. *Addictive Behaviors*. 2010; 27(6):867–886.
- Midanik LT, Greenfield TK, Bond J. Addiction sciences and its psychometrics: The measurement of alcohol-related problems. *Addiction*. 2007; 102(11):1701–1710. DOI: 10.1111/j.1360-0443.2007.01886.x [PubMed: 17610540]
- Miller, WR., Tonigan, JS., Longabaugh, R. *The Drinker Inventory of Consequences (DrInC): An Instrument for Assessing Adverse Consequences of Alcohol Abuse*. Vol. 4. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism; 1995. Test manual (Project MATCH Monograph Series
- Moos RH, Finney JW. The expanding scope of alcoholism treatment evaluation. *American Psychologist*. 1983; 38(10):1036–1044. DOI: 10.1037/0003-066X.38.10.1036 [PubMed: 6314859]
- Muthén, LK., Muthén, BO. *Mplus users guide (Version 7)*. 2012.
- Neale J, Finch E, Marsden J, Mitcheson L, Rose D, Strang J, ... Wykes T. How should we measure addiction recovery? Analysis of service provider perspectives using online Delphi groups. *Drugs: Education, Prevention and Policy*. 2014; 21(4):310–323. DOI: 10.3109/09687637.2014.918089
- Ohaeri JU, Awadalla AW, El-Abassi AHM, Jacob A. Confirmatory factor analytical study of the WHOQOL-Bref: experience with Sudanese general population and psychiatric samples. *BMC medical research methodology*. 2007; 7(1):37. doi: 10.1186/1471-2288-7-37 [PubMed: 17678527]
- Skevington SM, Lofly M, O'Connell KA. The World Health Organization's WHOQOL-BREF Quality of life assessment: Psychometric properties and results of the international field trial A report from the WHOQOL Group. *Quality of Life Research*. 2004; 13(2):299–310. [PubMed: 15085902]
- Skinner HA, Allen BA. Alcohol dependence syndrome: Measurement and validation. *Journal of Abnormal Psychology*. 1982; 91:199–209. [PubMed: 7096790]
- Tiffany ST, Friedman L, Greenfield SF, Hasin DS, Jackson R. Beyond drug use: a systematic consideration of other outcomes in evaluations of treatments for substance use disorders. *Addiction*. 2012; 107:709–718. [PubMed: 21981638]
- Tracy EM, Laudet A, Min MO, Kim H, Brown S, Jun M, Singer L. Prospective patterns and correlates of quality of life among women in substance abuse treatment. *Drug and Alcohol Dependence*. 2012; 124(3):242–249. DOI: 10.1016/j.drugalcdep.2012.01.010 [PubMed: 22333265]
- Trompenaars FJ, Masthoff ED, Van Heck GL, Hodiament PP, De Vries J. Content validity, construct validity, and reliability of the WHOQOL-Bref in a population of Dutch adult psychiatric outpatients. *Quality of Life Research*. 2005; 14(1):151–160. DOI: 10.1007/s11136-004-0787-x [PubMed: 15789949]
- Volk RJ, Cantor SB, Steinbauer JR, Cass AR. Alcohol use disorders, consumption patterns, and health-related quality of life in primary care patients. *Alcoholism: Clinical and Experimental Research*. 1997; 21:899–905.
- Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care*. 1992; 30:473–481. [PubMed: 1593914]
- WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*. 1998; 28:551–558. [PubMed: 9626712]

- Witkiewitz K. "Success" following alcohol treatment: moving beyond abstinence. *Alcoholism: Clinical and Experimental Research*. 2013; 37(S1):E9–E13. DOI: 10.1111/acer.12001
- World Health Organization. *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*. 2015.
- Yao G, Wu CH. Factorial invariance of the WHOQOL-BREF among disease groups. *Quality of Life Research*. 2005; 14(8):1881–1888. DOI: 10.1007/s11136-005-3867-7 [PubMed: 16155775]
- Zubaran C, Foresti K. Quality of life and substance use: concepts and recent tendencies. *Current Opinion in Psychiatry*. 2009; 22(3):281–286. DOI: 10.1097/YCO.0b013e328328d154 [PubMed: 19378380]

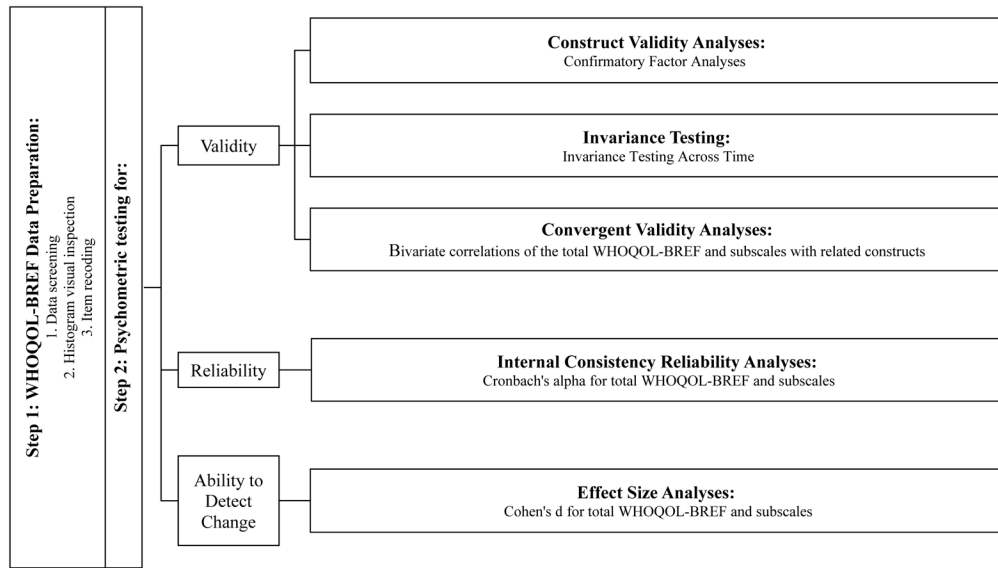


Figure 1. Data preparation and analysis plan

Note. Measure abbreviation used is WHOQOL-BREF (World Health Organization Quality of Life, Brief measure)

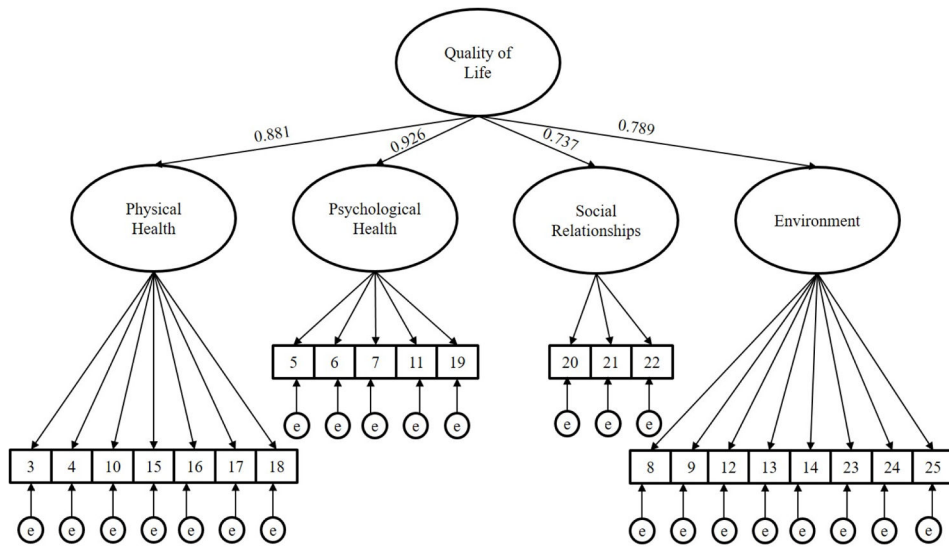


Figure 2. Confirmatory factor analysis structure of the WHOQOL-BREF in COMBINE at baseline with higher-order factor loadings (n=672, RMSEA=0.054 (90% CI: 0.050, 0.059), CFI=0.940, TLI=0.933; n=679, RMSEA=0.050 (90% CI: 0.045, 0.055), CFI=0.942, TLI=0.935)

Table 1

Standardized item loadings from final CFA with replication sample (n =679)

WHOQOL-BREF Item	Physical Health Factor	Psychological Health Factor	Social Relationships Factor	Environment Factor
3. Physical pain	0.450			
4. Medical treatment	0.306			
5. Enjoy life		0.784		
6. Life is meaningful		0.747		
7. Concentration		0.728		
8. Safety				0.765
9. Physical environment				0.714
10. Energy	0.795			
11. Bodily appearance		0.635		
12. Money				0.661
13. Availability of information				0.728
14. Leisure opportunities				0.523
15. Ability to get around	0.650			
16. Sleep	0.544			
17. Daily living activities	0.871			
18. Work capacity	0.852			
19. Personal abilities		0.756		
20. Personal relationships			0.875	
21. Sex life			0.641	
22. Friend support			0.635	
23. Living conditions				0.628
24. Access to health services				0.578
25. Transportation				0.577

Table 2

Comparison of CFA model fit between widely-cited published factor structures of the WHOQOL-BREF with COMBINE replication sample (n=679)

Citation	RMSEA (90% CI)	CFI	TLI
Jaracz et al., 2006	0.053 (0.048–0.058)	0.944	0.936
Skevington et al., 2004	0.050 (0.045–0.055)	0.942	0.935
Trompenaars et al., 2005	0.053 (0.048–0.058)	0.938	0.930
Yao & Wu, 2002	0.063 (0.059–0.068) *	0.908 *	0.897 *

Note:

* indicates non-positive definite matrix; results should be interpreted very cautiously

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

WHOQOL-BREF measurement invariance across time using the method described by Chen and colleagues (2005)

	RMSEA (90% CI)	CFI	TLI
Model 1: Baseline to Week 26 (<i>N</i> =1381)	0.037 (0.035–0.038)	0.921	0.916
Model 2: Baseline to Week 26 (<i>N</i> =1381)	0.035 (0.034–0.037)	0.926	0.923
Model 3: Baseline to Week 26 (<i>N</i> =1381)	0.035 (0.033–0.037)	0.927	0.924
Model 4: Baseline to Week 26 (<i>N</i> =1381)	0.033 (0.032–0.035)	0.927	0.931
Model 5: Baseline to Week 26 (<i>N</i> =1381)	0.035 (0.033–0.036)	0.920	0.925
Model 1: Week 26 to Week 52 (<i>N</i> =1123)	0.042 (0.040–0.044)	0.917	0.912
Model 2: Week 26 to Week 52 (<i>N</i> =1123)	0.040 (0.038–0.041)	0.924	0.921
Model 3: Week 26 to Week 52 (<i>N</i> =1123)	0.039 (0.038–0.041)	0.925	0.922
Model 4: Week 26 to Week 52 (<i>N</i> =1123)	0.039 (0.036–0.039)	0.925	0.929
Model 5: Week 26 to Week 52 (<i>N</i> =1123)	0.037 (0.036–0.039)	0.926	0.930

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Internal consistency reliability of the WHOQOL-BREF total scale and subscales as administered in COMBINE

	Baseline α	Week 26 α	Week 52 α
Full WHOQOL-BREF (items 3–25)	0.901	0.929	0.926
Physical Health Subscale (items 3, 4, 10, 15, 16, 17, 18)	0.768	0.819	0.816
Psychological Health Subscale (items 5, 6, 7, 11, 19)	0.770	0.837	0.821
Social Relationships Subscale (items 20, 21, 22)	0.718	0.761	0.746
Environmental Subscale (items 8, 9, 12, 13, 14, 23, 24, 25)	0.812	0.846	0.846

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Convergent validity of the WHOQOL-BREF tested via bivariate correlations

	Total Score	Physical Health Subscale	Psychological Health Subscale	Social Relationships Subscale	Environment Subscale
ADS Total	-0.371***	-0.336***	-0.004	-0.001	-0.009
BSI Global Severity	-0.683***	-0.601***	-0.045	-0.037	-0.055*
BSI Somatization	-0.508***	-0.541***	-0.064*	-0.058*	-0.074**
BSI Obsessive-Compulsiveness	-0.578***	-0.534***	-0.054*	-0.045	-0.058*
BSI Interpersonal Sensitivity	-0.592***	-0.493***	-0.022	-0.015	-0.030
BSI Depression	-0.665***	-0.522***	-0.039	-0.031	-0.047
BSI Anxiety	-0.534***	-0.496***	-0.048	-0.041	-0.057*
BSI Hostility	-0.458***	-0.404***	-0.041	-0.036	-0.049
BSI Phobic Anxiety	-0.507***	-0.441***	-0.031	-0.025	-0.042
BSI Paranoid Ideation	-0.510***	-0.405***	-0.038	-0.035	-0.053
BSI Psychoticism	-0.602***	-0.490***	-0.035	-0.029	-0.041
DrInC Total	-0.452***	-0.386***	-0.019	-0.016	-0.024
DrInC Physical Consequences Subscale	-0.412***	-0.411***	-0.011	-0.008	-0.013
DrInC Relationship Consequences Subscale	-0.315***	-0.253***	-0.027	-0.027	-0.033
DrInC Intrapersonal Consequences Subscale	-0.405***	-0.345***	-0.017	-0.014	-0.018
DrInC Impulsive Actions Subscale	-0.318***	-0.240***	0.007	0.009	0.000
DrInC Social Responsibilities Subscale	-0.460***	-0.386***	-0.023	-0.020	-0.032

Note.

* p < .05;

** p < .01;

*** p < .001

ADS = Alcohol Dependence Scale. BSI = Brief Symptom Inventory. DrInC = Drinker Inventory of Consequences

Table 6

Cohen's d effect sizes comparing baseline subscale and total WHOQOL-BREF summary scores with week-26 and week-52 timepoints within the full sample

	Baseline versus Week 26	Baseline versus Week 52
Physical Health Subscale	0.44	0.40
Psychological Health Subscale	0.44	0.49
Social Relationships Subscale	0.40	0.40
Environment Subscale	0.33	0.34
Total Summary Score	0.48	0.51

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7

Cohen's d effect sizes comparing Naltrexone versus Placebo and Combined Behavioral Intervention (CBI) versus Medication Management (MM) subscale and total WHOQOL-BREF summary scores with week-26 and week-52 timepoints

	Naltrexone versus Placebo	CBI versus MM
Physical Health Subscale, Week 26	0.06	0.06
Physical Health Subscale, Week 52	0.06	0.14
Psychological Health Subscale, Week 26	0.03	0.03
Psychological Health Subscale, Week 52	0.10	0.05
Social Relationships Subscale, Week 26	0.04	0.08
Social Relationships Subscale, Week 52	0.13	0.05
Environment Subscale, Week 26	0.09	0.08
Environment Subscale, Week 52	0.11	0.01
Total Summary Score, Week 26	0.07	<0.01
Total Summary Score, Week 52	0.11	0.07

Note. CBI = Combined Behavioral Intervention, MM = Medication Management

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript