



# HHS Public Access

Author manuscript

*Psychol Addict Behav.* Author manuscript; available in PMC 2019 February 01.

Published in final edited form as:

*Psychol Addict Behav.* 2018 February ; 32(1): 52–63. doi:10.1037/adb0000344.

## Revisiting the Drinker Inventory of Consequences: An Extensive Evaluation of Psychometric Properties in Two Alcohol Clinical Trials

**Megan Kirouac and Katie Witkiewitz**

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

### Abstract

Alcohol-related consequences are linked directly to the diagnostic criteria for alcohol use disorder (AUD). However, alcohol consumption outcome variables (e.g., percent days abstinent, heavy drinking days) remain the dominant outcome in AUD treatment research. Two reasons AUD treatment researchers have not shifted to include alcohol-related consequences as a primary outcome may be that previous studies have failed to provide convincing evidence of (1) the psychometric properties of measures of alcohol-related consequences, and (2) whether consequences measures are sensitive to change following treatment. The present study directly addresses these two concerns via psychometric evaluation and sensitivity/specificity testing of the Drinker Inventory of Consequences (DrInC; Miller, Tonigan, & Longabaugh, 1995) in two of the largest multisite clinical trials ever conducted (COMBINE Study, Anton et al., 2006; and Project MATCH, Project MATCH Research Group, 1997). Results indicated that the five subscales commonly used for the DrInC had poor construct validity and were non-invariant across time. A newly developed three factor model consisting of mild, moderate, and severe consequences had excellent psychometrics, including good internal consistency reliability, construct validity, and measurement invariance over time. The three factor model of the DrInC was also sensitive and specific for detecting consumption outcomes in both COMBINE and MATCH and had convergent validity with measures of consumption and wellbeing. In conclusion, the three factor DrInC may be a useful tool for defining AUD treatment success in a clinically meaningful way that aligns with diagnostic criteria.

### Keywords

Alcohol use disorder treatment; COMBINE Study; Project MATCH; alcohol-related consequences; treatment outcomes

### Introduction

Alcohol-related disorders have been part of the American Psychiatric Association (APA) Diagnostic and Statistical Manual (DSM) since it was first published in 1952 (APA, 1952).

---

Correspondence concerning this article should be addressed to: 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.

The authors report no conflicts of interest.

More recently, alcohol-related disorders were conceptualized as alcohol abuse and alcohol dependence in the DSM-IV-TR and most recently in DSM-5 on a spectrum of alcohol use disorder (AUD; APA, 2000; 2013). Based on DSM-5, AUD is currently conceptualized by symptoms of physical dependence (e.g., alcohol withdrawal and tolerance) and by the experience of alcohol-related consequences, such as recurrent failure to fulfill role obligations and giving up important activities due to alcohol use (APA, 2013). Such consequences lie at the core of how clients and their loved ones view alcohol use disorders and the recovery process (i.e., reduction in alcohol-related consequences; Kaskutas et al., 2014) and harm reduction treatment modalities (e.g., Marlatt & Witkiewitz, 2010; Midanik, Greenfield, & Bond, 2007; Witkiewitz, 2013).

Importantly, there is a disconnect between the aforementioned perspectives and AUD research: the primary outcomes examined in AUD treatment research are consumption-based outcomes (e.g., percent days abstinent, heavy drinking days; Falk et al., 2010). Two primary reasons consumption-based outcomes remain the dominant outcome in AUD treatment research include the lack of thorough psychometric vetting of non-consumption outcomes, such as measures of alcohol-related consequences, and it is generally assumed that alcohol-related consequence measures are likely to be insensitive to treatment effects. For instance, the Food and Drug Administration recently stated: “Trials intended to show direct effects on physical or psychosocial consequences of [alcohol] use...may be impractical” and stated that consumption variables must, therefore, be used as a “surrogate endpoint” (FDA, 2015, p. 2). However, this statement assumes insensitivity of alcohol-related consequence measures, even though sensitivity has never been tested empirically. The present study directly addresses these two concerns via a psychometric evaluation of the Drinker Inventory of Consequences (DrInC; Miller, Tonigan, & Longabaugh, 1995) in two of the largest multisite clinical trials ever conducted: the COMBINE Study (Anton et al., 2006) and Project MATCH (Project MATCH Research Group, 1997).

### **The Drinker Inventory of Consequences**

The Drinker Inventory of Consequences (DrInC; Miller et al., 1995) is a 45-item measure of alcohol-related consequences on which higher scores reflect greater alcohol consequences. The DrInC was initially conceptualized as containing five consequence subscales: Physical Consequences, Interpersonal Consequences, Intrapersonal Consequences, Impulse Control, and Social Responsibility (Miller et al., 1995). It was according to these five subscales that an abbreviated version of the DrInC was created: the Short Inventory of Problems (SIP; Feinn, Tennen, & Kranzler, 2003).

Numerous studies have explored the factor structure of the DrInC and the SIP via various methodologies and findings have been mixed. The five factor structure, based on the five consequence subscales, originating from early conceptualizations of the DrInC and SIP has been most widely examined, although support for this structure has been inconsistent. In the only psychometric study of the factor structure of the DrInC conducted to date, Forcehimes et al. (2007) failed to find evidence of the five factor model of the DrInC in Project MATCH when using confirmatory factor analyses (CFA) or exploratory factor analyses (EFA). Conversely, Kenna and colleagues (2005) reported “good” model fit per CFA of the five

factor model of the SIP (CFI=0.92; p. 435). Marra and colleagues (2014) found a one factor model (i.e., an overall alcohol-consequence severity factor) of the SIP, tested via CFA, provided the best fit to the data and was measurement invariant across Spanish and English speakers. Similarly, Alterman and colleagues (2009) examined the original five factor model of the SIP using CFA and found poor model fit; they argued exploratory factor analysis with varimax rotation indicated a one factor solution was best since the single factor explained 88.6% of the correlation structure of the SIP. Further, Bender and colleagues (2007) found evidence for the one factor solution using unrotated principal components analysis (PCA) of versions of the SIP used for Substance Use Disorder and Bipolar Disorders. Hagman et al. (2009) also concluded that a one factor solution of the SIP (adapted to assess alcohol and drug consequences) was supported per CFA analyses and subsequent Item Response Theory (IRT) analyses that were used to reduce the 15-item SIP to a 10-item version. However, Kiluk and colleagues (2013) tested a five factor, one factor, and higher-order factor models of the SIP using CFA and found the higher-order five factor model provided the best fit to their data, although the fit was not acceptable using rule-of-thumb conventions for determining acceptable model fit (Hu & Bentler, 1999). These alternative models were previously tested in the SIP and all failed to yield good model fit (Feinn et al., 2003).

One reason factor analytic studies of the DrInC and SIP have been so inconsistent may be due to the varying analytic methodologies (EFA, PCA, CFA, IRT) and the different versions of the SIP that have been administered across studies. The different versions of the SIP that have been administered in research is particularly troubling considering the SIP was developed on the assumption that the DrInC consisted of five factors, which has not been supported (e.g., Forcehimes et al., 2007). As such, it is crucial to extensively examine the underlying factor structure of the DrInC before researchers may proceed with creating an abbreviated version of the measure. Additionally, despite the fact that the DrInC and SIP are often used longitudinally, measurement invariance over time has never been examined for the DrInC or abbreviated versions of the DrInC. Examining measurement invariance of a factor model is critically important for assessing pre- to post-treatment change, which is often the goal of administrations of the DrInC in alcohol treatment research. Specifically, if measurement invariance over time is not supported (i.e., the factor structure changes over time) then changes in pre- and post-treatment scores on these measures may reflect changes in the measurement, rather than clinically meaningful changes in the construct itself.

Another possible explanation for the varying results of the factor structure for the DrInC and the SIP is that no prior studies have ever adjusted for clustering in the data. As discussed by Heck (2009), as well as Muthén and Muthén (2012), differences arising from un-measured variables may drive model fit in complex survey data (such as multisite AUD treatment research data); however, model fit can be improved by adjusting model estimates for differences across sites. Without adjustments for clustering, previous work may have been driven by within sample differences, which could explain the inconsistent (i.e., sample-specific) factor structures of the DrInC and the SIP. The present study controlled for treatment site differences via clustering and also attempted model replication between COMBINE and MATCH for improved confidence in the generalizability of the present findings.

## Methods

### Data

The present study used data collected from the COMBINE Study (Anton et al., 2006) and Project MATCH (Project MATCH Research Group, 1997). Table 1 summarizes the participant demographics, design, and exclusion criteria used in these two studies. COMBINE ( $N=1383$ ; Anton et al., 2006) was a large, multisite, randomized controlled trial of medications (acamprosate, naltrexone, or placebo equivalents) and psychosocial interventions (medication management or combined behavioral intervention) for individuals with alcohol dependence. Project MATCH ( $N=1726$ ; Project MATCH Research Group, 1997) was a large, multisite, randomized controlled trial of three psychosocial treatments (cognitive-behavioral treatment, motivation enhancement treatment, and twelve-step facilitation) for individuals with alcohol abuse or dependence. Participants were all seeking treatment and assessments included in the present analyses were conducted at baseline, post-treatment (4 months following baseline in COMBINE and 3 months following baseline in MATCH), and 12-months following the end of treatment. Participants in the COMBINE Study were more homogeneous and medically stable (a requirement for receiving study medications) than those recruited for Project MATCH, especially regarding alcohol problem severity.

### Measures

**Drinker Inventory of Consequences (DrInC)**—The DrInC was administered at multiple timepoints in both COMBINE and MATCH. For the present analyses, baseline, post-treatment, and 12-month follow-up DrInC data were analyzed. To examine the potential promise of the shorter version of the DrInC, the 15 items from the DrInC that comprise the SIP were analyzed separately in CFA.

In Project MATCH, the DrInC was non-uniformly administered to individuals who reported 100% days of abstinence during the follow-up assessments. Specifically, the items on the follow-up version of the DrInC are worded such that items should be endorsed only in reference to consequences that occurred due to drinking during the assessment window and some assessors in MATCH did not administer the DrInC to some, but not all, of the individuals who were abstinent at follow-up.

**Consumption Variables**—Both COMBINE and MATCH employed the Form 90 (Miller, 1996) to collect 90-day assessment window information on daily drinking levels. From these data, primary consumption outcome variables were computed: binary abstinence or any drinking (Abstinence), binary heavy drinking (HD), percent days abstinent (PDA), and percent heavy drinking days (PHDD). Standard drinks were calculated as 14 grams and “heavy drinking” was 4/5 or more standard drinks for women/men (NIAAA, 2004). Continuous consumption variables (PDA, PHDD) were used for convergent validity analyses of the DrInC and binary consumption variables (Abstinence, HD) were used to examine sensitivity/specificity of the DrInC via Receiver Operating Characteristic curves (ROC curves, described below; Hanley & McNeil, 1982).

**Convergent Validity Variables**—Convergent validity was examined via bivariate correlations of related measures at baseline timepoints. Specifically, correlations between the DrInC, alcohol consumption (Form 90 described above), alcohol craving measures, and wellbeing were examined. It was hypothesized the DrInC would be significantly, positively correlated to all consumption variables (except abstinence, which would be negatively correlated) and alcohol craving measures; conversely, significant negative correlations were hypothesized between the DrInC and wellbeing measures, AAI, and employment status/income.

For alcohol craving/temptation assessment, the Alcohol Abstinence Self-Efficacy Scale (AASE; DiClemente, Carbonari, Montgomery, & Hughes, 1994) was examined in COMBINE and MATCH; lower AASE scores indicated higher temptation/craving to drink. In addition to the AASE, an individual item assessing overall temptation/craving was administered in MATCH. In COMBINE, temptation/craving was also measured by the Obsessive-Compulsive Drinking Scale (OCDS; Anton, 2000), where higher scores indicated greater alcohol craving. Finally, the Alcoholics Anonymous Involvement scale was administered in MATCH (Tonigan, Connors, & Miller 1996), which assessed for attendance of AA meetings as well as involvement with each of the 12-steps of AA. Wellbeing was assessed in COMBINE via the World Health Organization Quality of Life, brief measure (WHOQOL-BREF; WHOQOL Group, 1998) and the Health Survey (SF-12; Ware, Kosinski, & Keller, 1996). Wellbeing in Project MATCH was assessed via the Psychosocial Functioning Inventory (PFI; Feragne, Longabaugh, & Stevenson, 1983).

A final metric for wellbeing used in COMBINE and MATCH consisted of items that assessed employment status and income (ESI). Only a single, categorical item was used for employment status in COMBINE and MATCH and income was assessed in COMBINE but not MATCH. Further, the employment status item had to be re-coded in COMBINE and MATCH to facilitate more meaningful categories for analyses (see Table 1).

## Data Analysis

Visual depiction of the analyses conducted for the present study are presented in Figure 1. All descriptive and sensitivity/specificity analyses were conducted in SPSS version 23 (IBM Corp, 2015); factor models were estimated in Mplus version 7.3 (Muthén & Muthén, 2012). Missing data in the latent factor models were handled with maximum likelihood estimation or mean-and-variance adjusted weighted least squares (WLSMV) estimation as recommended by Kline (2011) and described in detail below. The present analyses included examinations of the following: construct validity and measurement invariance across time (via confirmatory factor analyses in Mplus), convergent validity (via bivariate correlations in SPSS), internal consistency reliability (via Cronbach's alpha in SPSS), and effect sizes (via Cohen's *d* for baseline to post-treatment and baseline to 12-month follow-up). Because total sample data were used, effect sizes will be referred to as “change scores” to avoid confusion with implications for treatment effects.

Confirmatory factor analysis (CFA) in the COMBINE Study and Project MATCH used baseline data to maximize sample size. CFA analyses were guided initially by factor structures that have been previously published for the DrInC and the SIP (Alterman et al.,

2009; Bender et al., 2007; Forcehimes et al., 2007; Hagman et al., 2009; Kenna et al., 2005; Miller et al., 1995). Data screening was conducted via SPSS version 23 (IBM Corp, 2015) to examine potential problems with the data prior to all analyses (e.g., nonnormality and outliers; Jackson, Gillaspay, & Purc-Stephenson, 2009). Per recommendations by Floyd & Widaman (1995), random split-half designs were used to test and replicate factor structures. The first half of the sample was used to find a model with acceptable model fit (defined below); the second half was used to replicate the model in an independent sample. Data were split randomly via SPSS version 23 (IBM Corp, 2015). Moreover, non-independence of observations within treatment sites was accounted for via clustering by recruitment site (i.e., “clinical research units” were the location where treatment and recruitment occurred) in all CFA and measurement invariance analyses using a sandwich estimator to calculate the standard errors (Heck, 2009; Muthén & Muthén, 2012). Clustering by site also adjusted for clustering within sites within arms for the Project MATCH study, where outpatient sites and aftercare sites were coded as different clinical research units.

Hu and Bentler (1998) recommended evaluating CFA fit based on indices that have different properties such as incremental fit and residual-based fit. In the present study, model fit was examined via the comparative fit index (CFI), the Tucker-Lewis Index (TLI), and the root-mean-square error of approximation (RMSEA). Several researchers have recommended the CFI as an alternative to other fit indices such as the chi-square test of fit that are easily influenced by sample size (e.g., Floyd & Widaman, 1995). Although some have advised against the use of “rules of thumb” for model fit (e.g., Marsh, Hau, & Wen, 2004; Yuan, 2005), others have argued that *a priori* fit indices cutoffs are important to retain objectivity in model evaluation (Jackson et al., 2009). *A priori* cutoffs for the above fit indices were used, as informed by Hu and Bentler (1999) in order to minimize Type I and Type II error rates and reflect good model fit: CFI>0.95; TLI>0.95; RMSEA<0.06. Acceptable model fit *a priori* cutoffs were CFI>0.90; TLI>0.90; RMSEA<0.08. Fit indices outside of these cutoffs were deemed inadequate.

When an adequately fitting factor solution was found and replicated in independent split-half sub-samples, measurement invariance across time was tested by examining nested models between baseline and post-treatment datasets (post-treatment timepoints were: 4-month follow-up in COMBINE and three month follow-up in MATCH). Measurement invariance over time was tested for possible non-equivalence of measurement parameters (e.g., item intercepts, item loadings) over time (Widaman et al., 2010). Specific procedures to test longitudinal measurement invariance followed the recommendations of Vandenberg and Lance (2000) based on the results of their literature review. First, an omnibus test of the equality of covariance matrices across time was tested. Next, configural invariance was tested wherein the overall factor structure is tested as equivalent across time (Horn & McArdle, 1992). Then, metric invariance was tested by constraining the factor loadings to be equivalent across time (Horn & McArdle, 1992). Next, thresholds were constrained to equality across time to establish scalar invariance (i.e., “strong invariance”). Residual invariance (i.e., “strict invariance;” Widaman et al., 2010) was not tested because the items of the DrInC are categorical and all residuals were automatically constrained to 1 for model identification, thus residual were already constrained to equality (at 1) across time-points. Studying measurement invariance over time is critical for assuming the changes in scores of

the DrInC from baseline to follow-up reflect true changes in alcohol-related consequences and not changes in the measurement of the construct itself.

A final property of the DrInC that was examined was its sensitivity and specificity for detecting consumption outcomes via Receiver Operating Characteristic curve analyses. The ROC results were evaluated using the area under the curve (AUC) where measures with AUC=1 are considered perfectly sensitive/specific to detection and discrimination of the target outcome variable and AUC<0.500 are considered poor (Bradley & Longstaff, 2004). Generally, AUC values>0.650 are considered adequately sensitive/specific (Egger & Borg, 2016). Although AUC reflects an ability to both accurately detect and discriminate a target outcome variable, for parsimony of language, AUC results will be described using “detection” language throughout the manuscript.

All ROC analyses were conducted for the DrInC timepoint that immediately followed treatment in each study (4-month follow-up in COMBINE and 3-month follow-up in MATCH). These analyses examined how sensitive/specific each variable is at detecting binary consumption outcomes at two timepoints: 4- or 3-months post-baseline and 12-months post-treatment for: 1) abstinence versus any drinking, and 2) no heavy drinking days versus any heavy drinking days (Falk et al., 2010). These binary endpoints were selected because they are currently the two endpoints for alcohol clinical trials recommended by the FDA (FDA, 2015). Analyses were conducted separately in COMBINE data and MATCH data to examine the cross validation of findings and ROC curve analyses were conducted for total DrInC score and DrInC subscale scores (for subscales upheld via CFA and invariance testing).

## Results

Descriptive results and change scores (Cohen's *d*; Cohen, 1988) are presented in Table 2; COMBINE and MATCH samples differed with respect to both consumption and DrInC variables. In COMBINE, the overall DrInC average summary score was 47.61 (SD=20.42; *N*=1381) at baseline, 13.36 (SD=18.85; *N*=1098) at post-treatment, and 19.89 (SD=21.81; *N*=965) at 12-month follow-up. In contrast, for MATCH the overall DrInC average summary score was 52.63 (SD=23.32; *N*=1703) at baseline, 35.86 (SD=26.78; *N*=985) at post-treatment, and 27.50 (SD=24.70; *N*=789) at 12-month follow-up. The largest change scores in COMBINE occurred between baseline and post-treatment ( $d=1.735$ ), whereas the largest change scores in MATCH occurred between baseline and 12-month follow-up ( $d=1.057$ ). Similar patterns are observed in the commonly used subscales (Physical Health consequences, Interpersonal consequences, Intrapersonal consequences, Impulse Control, and Social Responsibility) of the DrInC for COMBINE and MATCH. Change scores were generally higher in COMBINE overall and substantively differed between COMBINE and MATCH in that the greatest changes in subscale scores occurred from baseline to post-treatment whereas the largest change scores in MATCH occurred baseline to 12-month follow-up. These changes may reflect the overall sample differences between COMBINE and MATCH. Importantly, change scores for PDA and PHDD were similar to those observed in the DrInC, especially for the COMBINE Study.

Although previously published factor structures were examined in the present study, including 1- and 5-factor models that have been previously examined for the DrInC and SIP (Alterman et al., 2009; Bender et al., 2007; Forchimes et al., 2007; Hagman et al., 2009; Kenna et al., 2005; Miller et al., 1995), none of these factor structures fit adequately in both COMBINE and MATCH while also being strongly invariant over time (see Table 3). No adequately fitting, invariant over time model was found for the SIP in the present study and thus we did not conduct additional psychometric analyses of the SIP, given lack of a good fitting model.

The only factor structure tested in the present study that yielded adequate fit in both COMBINE and MATCH and was strongly invariant across time was a three factor solution created in the present study based on conceptualization of the DrInC as consisting of alcohol-related consequences that occur at different rates of AUD severity. Specifically, these three factors may be conceptualized as consequences that occur at mild severity thresholds, such as hangovers; consequences that occur at more moderate thresholds, such as taking foolish risks; and severe consequences, such as being arrested for driving while intoxicated. This three factor solution fit adequately at baseline in the second split half samples (COMBINE: RMSEA=0.041 (90% CI: 0.038, 0.043); CFI=0.920; TLI=0.916; MATCH: RMSEA=0.040 (90% CI: 0.038, 0.042); CFI=0.908; TLI=0.904). When testing measurement invariance of this three factor model using the full samples from COMBINE and MATCH, fit improved as additional constraints were added through constraining thresholds to equivalence across time (COMBINE: RMSEA=0.024 (90% CI: 0.023, 0.025); CFI=0.951; TLI=0.952; MATCH: RMSEA=0.018 (90% CI: 0.017, 0.019); CFI=0.941; TLI=0.942).

This three factor model is presented in Figure 2; mean and standard deviations at baseline and post-treatment are listed in Table 2 for these three factors. The three factors were significantly correlated at both time-points in both COMBINE and MATCH (baseline correlations between factors are shown in Figure 2). Baseline to post-treatment change scores for these three factors were large in COMBINE and medium to large in MATCH (see Table 2). The largest change scores were for the Mild Consequences factor in both COMBINE and MATCH ( $d=-1.834$ ,  $d=-0.684$ ). The smallest change scores were for the Severe Consequences factor in both COMBINE and MATCH ( $d=-1.180$ ,  $d=-0.455$ ). The total DrInC and the 3 factors also all had strong internal consistency: total DrInC  $\alpha=0.937$  in COMBINE,  $\alpha=0.938$  in MATCH; Mild Consequences factor  $\alpha=0.855$  in COMBINE,  $\alpha=0.833$  in MATCH; Moderate Consequences factor  $\alpha=0.905$  in COMBINE,  $\alpha=0.905$  in MATCH; Severe Consequences factor  $\alpha=0.808$  in COMBINE,  $\alpha=0.830$  in MATCH).

Convergent validity of the DrInC total score and new three factor model, assessed at baseline, are detailed in Table 4; convergent validity of the original five factor model is presented in Table 5. For the COMBINE Study, total DrInC and each of the three factors were negatively correlated with PDA as predicted, although the correlation for the Mild Consequences factor was non-significant ( $p>0.05$ ). There was also a non-significant ( $p>0.05$ ) correlation with the Severe Consequences factor and PHDD in COMBINE. All other bivariate correlations in COMBINE with the total DrInC and three factor summary scores were significant and in the hypothesized direction, indicating overall good convergent



validity of the total DrInC and the three factors in COMBINE. In MATCH, PDA was significantly, negatively correlated with all but the Severe Consequences factor, which was non-significantly, negatively correlated with PDA. However, PHDD, PFI, and Employment in MATCH were all significantly correlated with the total DrInC summary score and the three-factor subscales as hypothesized. Unexpected results were found for the correlation with the single temptation/craving item in MATCH (the Mild Consequences factor was non-significantly correlated and all other DrInC variables were negatively correlated, counter to hypotheses) as well as the AAI, which was significantly, positively correlated with all DrInC variables (counter to hypotheses). Similar convergent validity patterns were found for the original 5-factor subscales of the DrInC (Table 5), which suggested convergent validity of the new 3-factor model is comparable to the total DrInC and the previously used five factors.

Different sensitivity/specificity results between the COMBINE Study and Project MATCH were also found. As detailed in Table 6, for COMBINE, the DrInC total summary score and three factors adequately detected post-treatment and 12-month follow-up abstinence and heavy drinking (AUCs>0.650), with the sole exception of AUC=0.645 for the Severe Consequences factor's ability to detect/discriminate 12-month follow-up abstinence. For Project MATCH, however, all AUC values were<0.650 except post-treatment heavy drinking was consistently adequately detected by DrInC total summary score and the three factor scores (AUCs>0.650).

## Discussion

The present study revisited the psychometrics of the Drinker Inventory of Consequences (DrInC) and Short Inventory of Problems (SIP) in two multisite alcohol clinical trials. Although the DrInC and the SIP have been evaluated in previous studies, the present study provides the most extensive evaluation of the DrInC conducted to date. Current findings supported the construct validity and measurement invariance across time for the DrInC using a new three factor model. Overall, the proposed three factor model of the DrInC had strong psychometric properties in both COMBINE and MATCH, as evidenced by good construct validity, strong measurement invariance over time, good internal consistency reliability, and good convergent validity with conceptually related measures. These three factors performed comparably to conceptualizations of the DrInC as a measure of overall consequences (i.e., a single-factor score), but had improved factor model fit and, unlike a 1-factor model, was invariant across time. Consequently, the total DrInC score may be useful in clinical practice, but potential measurement non-invariance of a single-factor model suggests a three-factor model may be more appropriate for longitudinal research designs that examine changes in drinking consequences over time.

The present results directly and empirically contradict the claim that measures of consequences are insensitive to change over time. Although there were differences in sensitivity/specificity analyses in COMBINE and MATCH, the total DrInC and the three factor subscales adequately detected and discriminated at least some consumption outcomes. Further, medium to large change scores for the DrInC total score and three factor subscales of the DrInC in COMBINE and MATCH indicate potential for AUD treatment interventions to impact alcohol-related consequences in a 12-month time-period. Largest change scores

were found for the baseline to post-treatment Mild Consequences factor scores, which may indicate mild consequences are more readily changed. Relatively smaller pre- to post-treatment change scores for the Moderate and Severe Consequences factors may also reflect that these items often persist after an individual has stopped or reduced their drinking (e.g., money problems, trouble with the law, relationships harmed). Work by Cisler and Zweben (1999) has suggested many items in the DrInC may be residual from previous alcohol use episodes (e.g., “Not had the life I want”) and may be endorsed misleadingly in a post-treatment window. Future research to abbreviate the DrInC may start by testing the psychometrics of the DrInC with only those items that are reasonably expected to change during the course of a treatment episode (e.g., “Hangover” and “Taken Foolish Risks”).

Additionally, the three-factor model makes conceptual sense and the factors provide clinically useful subscales for assessing the change in consequence severity over time. Specifically, clinicians may examine changes in mild, moderate, and severe consequences during the course of treatment, rather than relying solely on a reduction in total consequences or a cut-point of outcomes. Examining severity of consequences is consistent with calls for multidimensional models of treatment outcomes that consider alcohol consequences, quality of life, and other dimensions of import to the specific client and clinician (Kaskutas et al., 2014). The results are also consistent with recent research on the limitations to approaching alcohol-related consequences from a summative framework (Lane & Sher, 2014). Using IRT, Lane and Sher (2014) demonstrated that the current DSM-5 summative framework may mislead clinicians and researchers since not all consequences equally reflect AUD severity. For example, IRT results indicated that endorsement of tolerance, withdrawal, and efforts to cut down were far more prevalent and “easy” to endorse than the more “severe” criteria of giving up important activities due to drinking, role interference, and interpersonal problems (Lane & Sher, 2014). These more severe DSM-5 criteria are similar to items on the Moderate and Severe Consequences factors identified in the current study. The results from Lane and Sher (2014) also map onto the present findings of poor model fit for a one factor solution to the DrInC or SIP and may emphasize why it is important to consider *the degree of severity of consequences* reported, rather than simply *how many consequences* are reported. As such, the present three factor model that classifies *specific* alcohol consequences (DrInC items) into categories of severity (mild, moderate, severe) is consistent with the findings of Lane and Sher (2014) to underscore the importance of severity of consequences, as opposed to total number of consequences.

The present evidence for conceptualizing the DrInC as consisting of three factors of varying levels of severity of consequences is also counter to previous conceptualizations upon which the abbreviated version of the DrInC (the SIP) was created. The present findings may highlight why previous studies and the present attempts to identify the factor structure of the SIP have been problematic. Since the DrInC does not appear to be comprised of five factors, the development of the SIP may have been misinformed and new efforts may be undertaken to abbreviate the DrInC based on the presently described three factor model of the DrInC instead.

## Limitations

The present study limitations are primarily that findings are constrained to administration of the DrInC and SIP in the COMBINE Study and Project MATCH. The timepoints were not parallel between the two studies and administration methodology differed between studies. The DrInC was inconsistently administered to individuals who reported total abstinence during the assessment window in MATCH and was administered to all individuals in COMBINE, regardless of drinking status. These differences in assessment administration could explain why ROC curve and convergent validity results were inconsistent between the two studies. Additionally, findings from the present study may be limited by the fact that the full samples in COMBINE and MATCH were used; large samples may be why some findings were statically significant, despite small correlations. Findings from the present study should be replicated in additional study samples.

Another potential limitation to the present study is that full samples were used for COMBINE and MATCH, despite the different treatment conditions (e.g., therapy and medication combination conditions in COMBINE and both therapy conditions and aftercare versus outpatient treatment arms in MATCH). Demographic characteristics within each treatment condition and in COMBINE and MATCH were heterogeneous, which may explain some of the different results found between samples. However, two advantages of including full samples are increased generalizability to heterogeneous treatment and research settings as well as increased stability of models used in the latent variable modeling for the present study. Utilizing the full samples increased our participant to parameter ratio, which has been advocated as a method of assuring greater model stability (e.g., Gorsuch, 1983; Streiner, 1994). Further, collapsing the datasets to use full samples rather than treatment-arm subsamples allowed for random split half CFA testing, which allowed for testing and replicating CFA model structures for even greater stability of the present findings (Floyd & Widaman, 1995).

It is also noteworthy that both COMBINE and MATCH had samples of participants who were actively seeking treatment. Consequently, the present factor structure may not replicate in other populations. For example, the Severe Consequences factor has an item for DUI as well as an item for the experience of vomiting/becoming sick. Though these two items may appear disparate when thinking of the general population, they make conceptual sense considering a treatment-seeking population. For individuals seeking treatment, experiencing vomiting/sickness due to drinking may be something that has not been experienced since early in one's clinical course of AUD; as such, experiencing vomiting/sickness upon seeking treatment for AUD may be more indicative of someone with severe AUD and potential additional medical problems caused or exacerbated by alcohol use. Accordingly, both DUI and vomiting/sickness items would likely be perceived by treatment-seeking clients as severely harmful consequences of alcohol use. In another context, such as college student drinking, these two items may be perceived as dissimilar; researchers should be careful to use population-specific measures for alcohol-related consequences.

There was also no acceptable factor model identified for the Short Inventory of Problems (SIP). The 15 items in the SIP failed to yield an adequately fitting model that was strongly invariant across time in both COMBINE and MATCH. It is unclear if further work could be

done to refine the SIP or create a new, brief version of the DrInC that might reduce participant burden and costs associated with lengthy assessment batteries. For example, the National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) has developed a brief version of alcohol use, consequences, and expectancies with promising psychometric properties (Pilkonis et al., 2013).

Another limitation was that the present study assumed the DrInC consists of reflective constructs whereby the latent variables cause the indicators. In other words, the present analyses assumed alcohol-related consequences as a construct are amenable to CFA analyses by assuming consequences represent an underlying construct that predicts individual item responses. But it is possible that the DrInC is represents a formative construct, whereby the individual item responses predict the ultimate level of alcohol-related consequences (i.e., indicators cause the latent construct (Kenny, 2016)). Per discussion with one of the creators of the DrInC, it was developed from a conceptualization that individual items would influence the overall level of alcohol-related consequences (i.e., formative construct) but existing analyses have consistently used a reflective construct approach (W. R. Miller, personal communication, May 23, 2017). The issue of formative versus reflective constructs has yet to be explored with the DrInC, but the present approach is conventional for latent variable analysis and confirmatory factor analyses are often used to examine construct validity of a measure (Kenny, 2016).

Similarly, the DrInC and other measures of consequences are limited by the fact that not everyone may possibly endorse every item. For instance, items assessing for marital problems would not apply to individuals who are not married. These inherent limitations to the DrInC and many other measures of alcohol-related consequences may have detrimentally impacted some of the convergent validity correlations or other properties as an effect of unmeasured sources of variance. Latent variable modeling, such as a multidimensional outcome modeling approach, may be one possible way for future researchers to circumvent these inherent limitations since measurement error is modeled directly into the model equations.

Moreover, there is emerging evidence suggesting that participants may not perceive alcohol-related consequences as negative consequences. For instance, some research suggests that heavy drinking college students do not necessarily view researcher-generated “negative consequences” as wholly negative (Merrill, Read, & Barnett, 2013) and participant's subjective perspectives of the consequences in the DrInC were not assessed in the present study. Similarly, Future research is needed to assess how well the “alcohol-related problems” of the DrInC map onto clients' subjective experiences. Future research on drinking consequences could endeavor to develop a brief measure with strong psychometric properties, with consideration of consequences that could occur for all participants (e.g., do not focus on “marital problems”), and that also inquires about participants' perception of their experiences (e.g., not assuming that participants view the consequence as negative). Addressing these limitations of the DrInC is imperative for future research and likely requires the development of a new measure. Based on the factor structure of consequences supported in the current study we recommend focusing item development on a mix of mild,

moderate, and severe consequences, which would yield a measure of drinking consequences that covers a range of severity of consequences.

## Conclusions

Based on the present findings, the three factor model of the DrInC appears appropriate for use in longitudinal AUD treatment research. Not only did the DrInC demonstrate good psychometric qualities and measurement invariance over time using a new three factor solution, but the present findings refute the belief that measures of consequences are insensitive to change over time. Future researchers may use the presently-reported DrInC three factor subscales to examine the potential benefit of AUD treatments in ways that are more clinically meaningful than consumption-based outcomes alone.

## Acknowledgments

This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA; R01-AA022328; PI: Witkiewitz; R21-AA017137; PI: Witkiewitz; F31-AA024959; PI: Kirouac).

## References

- Alterman AI, Cacciola JS, Ivey MA, Habing B, Lynch KG. Reliability and validity of the alcohol short index of problems and a newly constructed drug Short Index of Problems. *Journal of Studies on Alcohol and Drugs*. 2009; 70(2):304. [PubMed: 19261243]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 1st. Washington, DC: Author; 1952.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed., text revision. Washington, DC: Author; 2000.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th. Washington, DC: Author; 2013.
- Anton RF. Obsessive-compulsive aspects of craving: development of the Obsessive Compulsive Drinking Scale. *Addiction*. 2000; 95(2)(January):S211–7. DOI: 10.1080/09652140050111771 [PubMed: 11002915]
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, et al. 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. DOI: 10.1001/jama.295.17.2003 [PubMed: 16670409]
- Bender RE, Griffin ML, Gallop RJ, Weiss RD. Assessing negative consequences in patients with substance use and bipolar disorders: Psychometric properties of the Short Inventory of Problems. *The American Journal on Addictions*. 2007; 16:503–509. DOI: 10.1080/10550490701641058 [PubMed: 18058418]
- Bradley AP, Longstaff ID. Sample size estimation using the receiver operating characteristic curve. *Proceedings of the 17th International Conference on Pattern Recognition*. 2004; 4:428–431. DOI: 10.1109/ICPR.2004.1333794
- Cisler RA, Zweben A. Development of a composite measure for assessing alcohol treatment outcome: operationalization and validation. *Alcoholism: Clinical and Experimental Research*. 1999; 23(2): 263–271. DOI: 10.1097/0000374-199902000-00011
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- DiClemente CC, Carbonari JP, Montgomery RP, Hughes SO. The Alcohol Abstinence Self-Efficacy scale. *Journal of Studies on Alcohol*. 1994; 55(2):141–148. DOI: 10.15288/jsa.1994.55.141 [PubMed: 8189734]

- Egger, D., Borg, JS. Introduction to binary classification. Mastering Data Analysis in Excel. 2016. [online lecture] Retrieved from Coursera Web Series by Duke University <https://www.coursera.org/learn/analytics-excel/lecture/TUihw/introduction-to-binary-classification>
- European Medicines Agency. Guideline on the development of medicinal products for the treatment of alcohol dependence. London: United Kingdom: 2010. p. 1-17.
- Falk D, Wan XQ, Liu L, Fertig J, Mattson M, Ryan M, et al. 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
- Feinn R, Tennen H, Kranzler HR. Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. *Alcoholism, Clinical and Experimental Research*. 2003; 27(9): 1436–41. DOI: 10.1097/01.ALC.0000087582.44674.AF
- Feragne MA, Longabaugh R, Stevenson JF. The Psychosocial Functioning Inventory. *Evaluation & The Health Professions*. 1983; 6(1):25–48. DOI: 10.1177/016327878300600102 [PubMed: 10259949]
- Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*. 1995; 7(3):286–299.
- Food and Drug Administration. Medical Review of Vivitrol: 21-897. U.S. Government; Rockville, MD: 2006.
- Food and Drug Administration. Alcoholism: Developing Drugs for Treatment: Guidance for Industry: 1-14. U.S. Government; Rockville, MD: 2015.
- Forcehimes AA, Tonigan JS, Miller WR, Kenna GA, Baer JS. Psychometrics of the Drinker Inventory of Consequences (DrInC). *Addictive Behaviors*. 2007; 32(8):1699–1704. DOI: 10.1016/j.addbeh.2006.11.009 [PubMed: 17182194]
- Gorsuch, RL. Factor Analysis. 2nd. Hillsdale, NJ: Erlbaum; 1983.
- Hagman BT, Kuerbis AN, Morgenstern J, Bux DA, Parsons JT, Heidinger BE. An Item Response Theory (IRT) analysis of the Short Inventory of Problems-Alcohol and Drugs (SIP-AD) among non-treatment seeking men-who-have-sex-with-men: Evidence for a shortened 10-item SIP-AD. *Addictive Behaviors*. 2009; 34(11):948–954. DOI: 10.1016/j.addbeh.2009.06.004 [PubMed: 19564078]
- Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*. 1982; 143:29–36. DOI: 10.1148/radiology.143.1.7063747 [PubMed: 7063747]
- Heck, RH., Thomas, SL. An Introduction to Multilevel Modeling Techniques. 2nd. New York, NY: Routledge; 2009.
- Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*. 1992; 18(3-4):117–144. DOI: 10.1080/03610739208253916 [PubMed: 1459160]
- 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. DOI: 10.1080/10705519909540118
- IBM Corp. IBM SPSS Statistics for Windows. Vol. Version 23.0. Armonk, NY: IBM Corp; 2015.
- Jackson DL, Gillaspay JR, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*. 2009; 14(1):6–23. DOI: 10.1037/a0014694 [PubMed: 19271845]
- Kaskutas LA, Borkman TJ, Laudet A, Ritter LA, Witdbrodt J, Subbaraman MS, et al. Bond J. Elements that define recovery: The experiential perspective. *Journal of Studies on Alcohol and Drugs*. 2014; 75:999–1010. DOI: 10.15288/jsad.2014.75.999 [PubMed: 25343658]
- Kenna GA, Longabaugh R, Gogineni A, Woolard RH, Nirenberg TD, Becker B, et al. Karolczuk K. Can the short index of problems (SIP) be improved? Validity and reliability of the three-month SIP in an emergency department sample. *Journal of Studies on Alcohol*. 2005; 66(3):433–437. DOI: 10.15288/jsa.2005.66.433 [PubMed: 16047535]
- Kenny, DA. Miscellaneous variables: Formative and second-order factors. 2016. Retrieved from <http://davidakenny.net/cm/mvar.htm>

- Kline, RB. Principles and practice of Structural Equation Modeling. 3rd. New York: Guilford Press; 2011.
- Lane SP, Sher KJ. Limits of current approaches to diagnosis severity based on criterion counts: An example with DSM-5 Alcohol Use Disorder. *Clinical Psychological Science*. 2014; doi: 10.1177/2167702614553026
- Lenhard, W., Lenhard, A. Dettelbach (Germany): *Psychometrica*; 2016. Calculation of Effect Sizes. available: [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html)
- Marra LB, Field CA, Caetano R, von Sternberg K. Construct validity of the Short Inventory of Problems among Spanish speaking Hispanics. *Addictive Behaviors*. 2014; 39:205–210. DOI: 10.1016/j.addbeh.2013.09.02 [PubMed: 24148139]
- Marsh HW, Hua KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*. 2004; 11(3):320–341. DOI: 10.1207/s15328007sem1103\_2
- Merrill JE, Read JP, Barnett NP. The way one thinks affects the way one drinks: subjective evaluations of alcohol consequences predict subsequent change in drinking behavior. *Psychology of Addictive Behaviors*. 2013; 27(1):42–51. DOI: 10.1037/a0029898 [PubMed: 22985054]
- Miller, WR. NIAAA Project MATCH Monograph Series. Vol. 5. Washington: Government Printing Office; 1996. Form 90: A structured assessment interview for drinking and related behaviors. NIH Publication No. 96-4004
- Miller, WR., Tonigan, JS., Longabaugh, R. The Drinker Inventory of Consequences (DrInC): An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1995.
- Muthén LK, Muthén BO. Mplus users guide. 2012; Version 7
- National Institute on Alcohol Abuse and Alcoholism. NIAAA Council approves definition of binge drinking. NIAAA Newsletter. 2004. Retrieved from: [http://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter\\_Number3.pdf](http://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.pdf)
- Pilkonis PA, Yu L, Colditz J, Dodds N, Johnston KL, Maihoefer C, et al. McCarty D. Item banks for alcohol use from the patient-reported outcomes measurement information system (PROMIS): Use, consequences, and expectancies. *Drug and Alcohol Dependence*. 2012; 130(0):167–177. DOI: 10.1016/j.drugalcdep.2012.11.002 [PubMed: 23206377]
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: project MATCH post-treatment drinking outcomes. *Journal of Studies on Alcohol*. 1997; 58:7–29. [PubMed: 8979210]
- Streiner DL. Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*. 1994; 39:135–140. DOI: 10.1177/070674379403900303 [PubMed: 8033017]
- Tonigan JS, Connors GJ, Miller WR. Alcoholics Anonymous Involvement (AAI) scale: Reliability and norms. *Psychology of Addictive Behaviors*. 1996; 10:75–80. DOI: 10.1037/0893-164X.10.2.75
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–69. DOI: 10.1177/109442810031002
- Ware JE Jr, Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*. 1996; 34:220–233. DOI: 10.1097/00005650-199603000-00003 [PubMed: 8628042]
- WHOQOL Group. Development of the World health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*. 1998; 28:551–558. [PubMed: 9626712]
- Widaman KF, Ferrer E, Conger RD. Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Development Perspectives*. 2010; 4:10–18. DOI: 10.1111/j.1750-8606.2009.00110.x [PubMed: 20369028]
- Witkiewitz K. Temptation to drink as a predictor of drinking outcomes following psychosocial treatment for alcohol dependence. *Alcoholism: Clinical and Experimental Research*. 2013; 37(3): 529–537. DOI: 10.1111/j.1530-0277.2012.01950
- World Health Organization. International Guide for monitoring alcohol consumption and related harm. Geneva, Switzerland: Author; 2000.

Yuan KH. Fit indices versus test statistics. *Multivariate Behavioral Research*. 2005; 40:115–148. DOI: 10.1207/s15327906mbr4001\_5 [PubMed: 26822275]

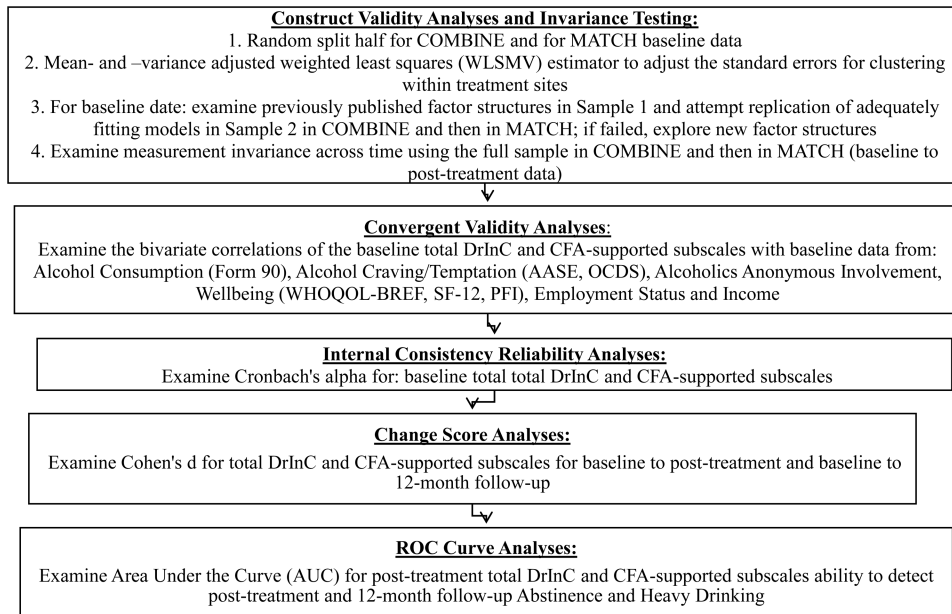
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





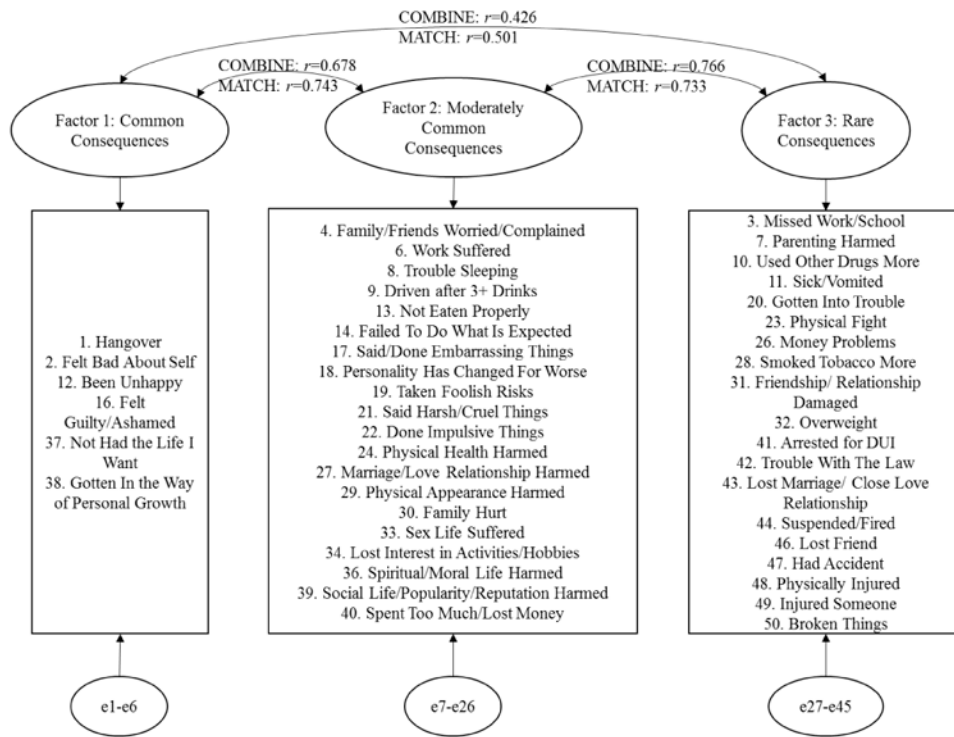
**Figure 1. Data analysis overview**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2. Factor structure for three factor model in COMBINE and MATCH with correlations between latent factors at baseline**

**Table 1**  
**Demographic, design, and exclusion criteria for COMBINE and MATCH**

	COMBINE	MATCH
Demographic characteristic		
Sample size	1383	1726
Gender -- % Male	69.1%	75.7%
Age – Mean (SD)	44.4 (10.2)	40.2 (10.9)
Ethnicity -- % White	76.8%	80.0%
Marital status -- % Married, in relationship	46.3%	41.4%
Employment status -- % Full or part-time	71.4%	82.1%
Higher education or equivalent	70.6%	53.4%
Design		
Randomization to treatment	9 groups	3 groups
Length of treatment	16 weeks	12 weeks
Follow-up assessments	12 months	12 months
Exclusion criteria		
Age	18+	18+
Meet criteria for abuse/dependence	Past year	Past year
Reading level	Literate	6 <sup>th</sup> grade
Comorbid psychiatric diagnoses	X	X
Unable to identify collateral informant		X
Severe cognitive impairment		X
Residential instability		X
Other illicit drug dependence	X	X

Note. COMBINE and MATCH employment items were recoded to represent increasing levels of employment: unemployment or disabled=0; homemaker, part-time employed, or retired=1; and full-time employed=2. COMBINE also included one item for income that was not paralleled in MATCH (<\$15,000; \$15,000-\$29,999; \$30,000-\$59,000; \$60,000-\$89,000; >\$90,000).

**Table 2**  
**Descriptive statistics and Cohen's d change scores for measures used in the present study at baseline, post-treatment (post-tx), and 12-month follow-up (12mo)**

	<i>N</i>	<i>M (SD)</i>	Cohen's <i>d</i>
Percent Days Abstinent: COMBINE Study	Baseline:	1383 21.41 (22.50)	Baseline to Post-tx: 1.809
	Post-Tx:	1288 72.66 (33.49)	
	12mo:	1099 62.63 (39.12)	Baseline to 12mo: 1.331
Percent Days Abstinent: Project MATCH	Baseline:	1725 30.90 (29.96)	Baseline to Post-tx: 1.786
	Post-Tx:	1657 83.17 (28.51)	
	12mo:	1594 76.69 (33.55)	Baseline to 12mo: 1.443
Percent Heavy Drinking Days: COMBINE Study	Baseline:	1383 70.52 (26.57)	Baseline to Post-tx: 1.919
	Post-Tx:	1288 17.54 (28.69)	
	12mo:	1171 26.20 (34.27)	Baseline to 12mo: 1.461
Percent Heavy Drinking Days: Project MATCH	Baseline:	1725 63.18 (31.43)	Baseline to Post-tx: 1.780
	Post-Tx:	1657 12.46 (25.09)	
	12mo:	1594 16.71 (29.17)	Baseline to 12mo: 1.530
DfnC: COMBINE Study	Baseline:	1381 47.61 (20.42)	Baseline to Post-tx: 1.735
	Post-Tx:	1098 13.36 (18.85)	
	12mo:	965 19.89 (21.81)	Baseline to 12mo: 1.320
Physical Health Subscale	Baseline:	1381 9.28 (4.36)	Baseline to Post-tx: 1.607
	Post-Tx:	1098 2.61 (3.87)	
	12mo:	965 3.95 (4.53)	Baseline to 12mo: 1.203
Interpersonal Subscale	Baseline:	1381 10.06 (6.01)	Baseline to Post-tx: 1.389
	Post-Tx:	1098 2.60 (4.44)	
	12mo:	965 4.06 (5.25)	Baseline to 12mo: 1.051
Intrapersonal Subscale	Baseline:	1381 14.44 (5.66)	Baseline to Post-tx: 1.749
	Post-Tx:	1098 4.45 (5.78)	
	12mo:	965 6.26 (6.66)	Baseline to 12mo: 1.343
Impulse Subscale	Baseline:	1381 7.56 (4.25)	Baseline to Post-tx: 1.414
	Post-Tx:	1098 2.11 (3.29)	

	<i>N</i>	<i>M (SD)</i>	<i>Cohen's d</i>
	12mo:	965	3.24 (3.83)
	Baseline to 12mo:		1.058
Social Responsibility Subscale	Baseline:	1381	6.27 (4.15)
	Post-Tx:	1098	1.60 (3.04)
	Baseline to Post-tx:		1.260
	12mo:	965	2.39 (3.46)
	Baseline to 12mo:		1.000
Mild Consequences Factor	Baseline:	1381	12.15 (4.30)
	Post-Tx:	1098	3.93 (4.70)
	Baseline to Post-tx:		-1.834
Moderate Consequences Factor	Baseline:	1381	25.46 (11.61)
	Post-Tx:	1098	6.63 (10.26)
	Baseline to Post-tx:		-1.707
Severe Consequences Factor	Baseline:	1381	10.00 (6.91)
	Post-Tx:	1099	2.79 (4.92)
	Baseline to Post-tx:		-1.1.80
DrInC: Project MATCH			
	Baseline:	1703	52.63 (23.32)
	Post-Tx:	985	35.86 (26.78)
	Baseline to Post-tx:		0.680
	12mo:	789	27.50 (24.70)
	Baseline to 12mo:		1.057
Physical Health Subscale	Baseline:	1626	9.48 (4.94)
	Post-Tx:	966	6.14 (5.13)
	Baseline to Post-tx:		0.666
	12mo:	818	5.12 (4.85)
	Baseline to 12mo:		0.945
Interpersonal Subscale	Baseline:	1558	12.21 (6.98)
	Post-Tx:	942	8.17 (7.34)
	Baseline to Post-tx:		0.568
	12mo:	807	6.09 (6.72)
	Baseline to 12mo:		0.888
Intrapersonal Subscale	Baseline:	1626	14.51 (6.02)
	Post-Tx:	964	10.38 (6.81)
	Baseline to Post-tx:		0.653
	12mo:	819	8.31 (7.16)
	Baseline to 12mo:		0.965
Impulse Subscale	Baseline:	1572	8.69 (5.10)
	Post-Tx:	967	6.06 (5.43)
	Baseline to Post-tx:		0.503
	12mo:	820	3.41 (4.21)
	Baseline to 12mo:		1.097
Social Responsibility Subscale	Baseline:	1598	7.49 (4.71)
	Post-Tx:	964	4.71 (4.69)
	Baseline to Post-tx:		0.591
	12mo:	822	4.73 (4.56)
	Baseline to 12mo:		0.592
Mild Consequences Factor	Baseline:	1714	11.49 (4.53)
	Baseline to Post-tx:		-0.684

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	<i>N</i>	<i>M (SD)</i>	Cohen's <i>d</i>
Post-Tx:	986	8.27 (5.00)	
Moderate Consequences Factor			
Baseline:	1716	27.04 (12.53)	Baseline to Post-tx: -0.675
Post-Tx:	986	18.17 (14.14)	
Severe Consequences Factor			
Baseline:	1717	13.30 (8.80)	Baseline to Post-tx: -0.455
Post-Tx:	986	9.20 (9.39)	

**Table 3**

**Model results for CFA and measurement invariance testing**

Measure (Dataset)	CFA Model	Invariance Testing Model	RMSEA (90% CI)	CFI	TLI
DrInC (COMBINE)	5-factors based on original conceptualization*		0.044 (0.041, 0.046)	0.900	0.894
	1-factor based on previously published models*		0.051 (0.049, 0.054)	0.861	0.854
	3-factor solution based on conceptualization of the DrInC as comprised of consequences that occur mild, moderate and severe	Configural: Baseline to Post-Treatment Loadings Constrained: Baseline to Post-Treatment Thresholds Constrained: Baseline to Post-Treatment	0.041 (0.038, 0.043)	0.920	0.916
DrInC (MATCH)	3-factor solution based on conceptualization of the DrInC as comprised of consequences that occur mild, moderate, and severe	Configural: Baseline to Post-Treatment Loadings Constrained: Baseline to Post-Treatment Thresholds Constrained: Baseline to Post-Treatment	0.017 (0.016, 0.019)	0.975	0.974
			0.019 (0.018, 0.020)	0.969	0.968
			0.024 (0.023, 0.025)	0.951	0.952
SIP (COMBINE)	5-factors based on original conceptualization	Configural: Baseline to Post-Treatment Loadings Constrained: Baseline to Post-Treatment Thresholds Constrained: Baseline to Post-Treatment	0.040 (0.038, 0.042)	0.908	0.904
			0.018 (0.017, 0.019)	0.945	0.944
			0.018 (0.017, 0.018)	0.946	0.946
SIP (MATCH)	1-factor based on previously published models*	Configural: Baseline to Post-Treatment Loadings Constrained: Baseline to Post-Treatment Thresholds Constrained: Baseline to Post-Treatment	0.018 (0.017, 0.019)	0.941	0.942
	5-factors based on original conceptualization	Configural: Baseline to Post-Treatment Loadings Constrained: Baseline to Post-Treatment Thresholds Constrained: Baseline to Post-Treatment	0.061 (0.053, 0.069)	0.969	0.960
			0.086 (0.084, 0.086)	0.894	0.883
		N/A (Not tested due to failed configural invariance)	N/A	N/A	
		N/A (Not tested due to failed configural invariance)	N/A	N/A	
		0.109 (0.102, 0.116)	0.890	0.871	
		0.077 (0.070, 0.084)	0.949	0.933	
		0.059 (0.057, 0.061)	0.894	0.883	
		N/A (Not tested due to failed configural invariance)	N/A	N/A	
		N/A (Not tested due to failed configural invariance)	N/A	N/A	

**Table 4**  
**Baseline measure convergent validity tested via bivariate correlations**

	<b>Total DrInC Summary Score</b>	<b>DrInC Mild Consequences Factor Subscale Summary Score</b>	<b>DrInC Moderate Consequences Factor Subscale Summary Score</b>	<b>DrInC Severe Consequences Factor Subscale Summary Score</b>
COMBINE PDA	$r = 0.061^*$	$r = -0.010$	$r = 0.060^*$	$r = 0.090^{**}$
COMBINE PHDD	$r = 0.091^{**}$	$r = 0.116^{***}$	$r = 0.091^{**}$	$r = 0.038$
COMBINE AASE	$r = 0.153^{***}$	$r = 0.116^{***}$	$r = 0.144^{***}$	$r = 0.146^{***}$
COMBINE OCDS	$r = 0.519^{***}$	$r = 0.473^{***}$	$r = 0.506^{***}$	$r = 0.386^{***}$
COMBINE WHOQOL-BREF	$r = -0.456^{***}$	$r = -0.361^{***}$	$r = -0.417^{***}$	$r = -0.413^{***}$
COMBINE SF-12	$r = -0.486^{***}$	$r = -0.388^{***}$	$r = -0.468^{***}$	$r = -0.397^{***}$
COMBINE Employment	$r = 0.225^{***}$	$r = -0.101^{***}$	$r = -0.202^{***}$	$r = -0.266^{***}$
COMBINE Income	$r = -0.233^{***}$	$r = -0.053$	$r = -0.187^{***}$	$r = -0.346^{**}$
MATCH PDA	$r = -0.136^{***}$	$r = -0.144^{***}$	$r = -0.176^{***}$	$r = -0.019$
MATCH PHDD	$r = 0.230^{***}$	$r = 0.218^{***}$	$r = 0.266^{***}$	$r = 0.094^{***}$
MATCH PFI	$r = -0.479^{***}$	$r = -0.396^{***}$	$r = -0.444^{***}$	$r = -0.409^{***}$
MATCH Employment	$r = 0.164^{***}$	$r = -0.068^{**}$	$r = -0.147^{***}$	$r = -0.175^{***}$
MATCH Temptation/Craving Item	$r = -0.060^*$	$r = 0.035$	$r = -0.069^{**}$	$r = -0.066^*$
MATCH AA Involvement	$r = 0.336^{***}$	$r = 0.277^{***}$	$r = 0.329^{***}$	$r = 0.261^{***}$

Note.

\*  
 $p < .05$ ;

\*\*  
 $p < .01$ ;

\*\*\*  
 $p < .001$ .

PDA = Percent Days Abstinent; PHDD = Percent Heavy Drinking Days; OCDS = Obsessive Compulsive Drinking Scale; AASE = Alcohol Abstinence Self-Efficacy Scale; WHOQOL-BREF = World Health Organization Quality of Life, Brief measure; SF-12 = Health Survey, 12-item; PFI = Psychosocial Functioning Inventory; AA Involvement = Alcoholics Anonymous Involvement Scale



**Table 5**  
**Baseline measure convergent validity tested via bivariate correlations for the original 5-factors of the DrInC**

	DrInC Physical Health Factor Subscale Summary Score	DrInC Interpersonal Factor Subscale Summary Score	DrInC Intrapersonal Factor Subscale Summary Score	DrInC Impulse Factor Subscale Summary Score	DrInC Social Responsibility Factor Subscale Summary Score
COMBINE PDA	$r = 0.000$	$r = 0.095^{***}$	$r = 0.006$	$r = 0.073^{**}$	$r = 0.092^{**}$
COMBINE PHDD	$r = 0.135^{***}$	$r = 0.040$	$r = 0.107^{***}$	$r = 0.041$	$r = 0.056^*$
COMBINE AASE	$r = 0.126^{***}$	$r = 0.144^{***}$	$r = 0.115^{***}$	$r = 0.137^{***}$	$r = 0.124^{***}$
COMBINE OCDS	$r = 0.462^{***}$	$r = 0.416^{***}$	$r = 0.499^{***}$	$r = 0.341^{***}$	$r = 0.460^{***}$
COMBINE WHOQOL-BREF	$r = -0.415^{***}$	$r = -0.320^{***}$	$r = -0.407^{***}$	$r = -0.321^{***}$	$r = -0.462^{**}$
COMBINE SF-12	$r = -0.476^{***}$	$r = -0.360^{***}$	$r = -0.443^{***}$	$r = -0.323^{***}$	$r = -0.444^{***}$
COMBINE Employment	$r = -0.199^{***}$	$r = -0.208^{***}$	$r = -0.136^{***}$	$r = -0.129^{***}$	$r = -0.279^{***}$
COMBINE Income	$r = -0.154^{***}$	$r = -0.179^{***}$	$r = -0.092^{**}$	$r = -0.221^{***}$	$r = -0.346^{***}$
MATCH PDA	$r = -0.225^{***}$	$r = -0.063^{**}$	$r = -0.149^{***}$	$r = -0.070^{**}$	$r = -0.073^{**}$
MATCH PHDD	$r = 0.307^{***}$	$r = 0.139^{***}$	$r = 0.229^{***}$	$r = 0.138^{***}$	$r = 0.162^{***}$
MATCH PFI	$r = -0.398^{***}$	$r = -0.413^{***}$	$r = -0.429^{***}$	$r = -0.310^{***}$	$r = -0.451^{***}$
MATCH Employment	$r = -0.144^{***}$	$r = -0.166^{***}$	$r = -0.099^{***}$	$r = -0.077^{**}$	$r = -0.196^{***}$
MATCH Temptation/Craving Item	$r = -0.054^*$	$r = -0.062^*$	$r = -0.013$	$r = -0.032$	$r = -0.083^{**}$
MATCH AA Involvement	$r = 0.218^{***}$	$r = 0.310^{***}$	$r = 0.331^{***}$	$r = 0.151^{***}$	$r = 0.322^{***}$

Note.

\*  $p < .05$ ;

\*\*  $p < .01$ ;

\*\*\*  $p < .001$ .

PDA = Percent Days Abstinent; PHDD = Percent Heavy Drinking Days; OCDS = Obsessive Compulsive Drinking Scale; AASE = Alcohol Abstinence Self-Efficacy Scale; WHOQOL-BREF = World Health Organization Quality of Life, Brief measure; SF-12 = Health Survey, 12-item; PFI = Psychosocial Functioning Inventory; AA Involvement = Alcoholics Anonymous Involvement Scale

**Table 6**  
**Receiver operating characteristic curve area under the curve (AUC) results for the**  
**Drinker Inventory of Consequences (DrInC) for detecting/discriminating post-treatment**  
**(post-tx) and 12-month follow-up (12mo) consumption outcomes in the COMBINE Study**  
**and Project MATCH**

	<b>DrInC total score</b>	<b>Factor 1 (Mild Consequences)</b>	<b>Factor 2 (Moderate Consequences)</b>	<b>Factor 3 (Severe Consequences)</b>
COMBINE Post-tx Abstinence	0.845	0.833	0.803	0.780
COMBINE Post-tx Heavy Drinking	0.845	0.840	0.824	0.782
COMBINE 12mo Abstinence	0.684	0.671	0.659	0.645
COMBINE 12mo Heavy Drinking	0.702	0.685	0.683	0.674
MATCH Post-tx Abstinence	0.583	0.585	0.573	0.586
MATCH Post-tx Heavy Drinking	0.679	0.672	0.673	0.671
MATCH 12mo Abstinence	0.424	0.432	0.425	0.436
MATCH 12mo Heavy Drinking	0.511	0.515	0.511	0.509