Deriving Alternative Criteria Sets for Alcohol Use Disorders Using Statistical Optimization: Results from the National Survey on Drug Use and Health

Cassandra L. Boness, Jordan E. Stevens, Douglas Steinley,

Timothy Trull, & Kenneth J. Sher

University of Missouri

Accepted Version* (11/1/2018)

*May differ from printed version

Experimental and Clinical Psychopharmacology

Author Note

Cassandra L. Boness, Jordan E. Stevens, Douglas Steinley, Timothy Trull, and Kenneth J. Sher, Department of Psychological Science, University of Missouri. The present study was supported by the NIH grants K05AA017242, T32AA013526, R01AA024133, and F31AA026177.

Correspondence should be addressed to Cassandra Boness, Department of Psychological Science, Psychology Building, 200 South Seventh Street, Columbia, MO 65211. *Phone*: +1 573 884 1485. *Fax*: +1 573 884 5588. *E-mail*: clmkdb@mail.missouri.edu.

These findings have been presented at the Research Society of Alcoholism annual conference. Otherwise, the ideas and findings have not been disseminated in any other form.

Abstract

Much of the foundation of clinical practice, psychiatric epidemiology, and research into the etiology, course, prevention, and treatment of alcohol use disorder (AUD) rests on psychiatric diagnosis. However, existing research has failed to adequately exploit empirical techniques and existing databases to derive criteria considered optimal with respect to predicting external correlates. The current project adopts a novel approach to deriving new diagnostic criteria sets and rules for AUD. Utilizing the 2010 (N = 24,120) and 2013 (N = 23,627) National Survey on Drug Use and Health (NSDUH; SAMHSA, 2011, 2014) datasets, we performed a statistical optimization procedure, using complete enumeration, on participants 21 or older who had consumed at least one alcoholic beverage in the past year. The goal was to maximize the distance (based on Cohen's d) between mean levels of the optimization criteria (i.e., consumption and functional impairment) in those with an AUD diagnosis versus those without. In contrast with current convention, AUD is derived transparently using a data-driven approach. The best solution included nine criteria with a diagnostic threshold of three, while the second-best solution comprised five criteria with a threshold of two. External validation demonstrated both solutions perform similarly, suggesting it is appropriate to use either, depending on the goal of the diagnosis. Overall, statistical optimization approaches can yield highly efficient criteria sets and rules although multiple, near equivalently performing solutions can be generated.

Keywords: alcohol use disorder, diagnosis, assessment

Public Significance Statement: This study suggests statistical methods can be used to derive highly efficient Alcohol Use Disorder (AUD) diagnoses, reducing the total number of diagnostic criteria needed to diagnose an individual with AUD while at the same time maintaining the

diagnosis' ability to predict relevant correlates such as treatment usage and co-morbid psychopathologies.

Deriving Alternative Criteria Sets for Alcohol Use Disorders Using Statistical Optimization: Results from the National Survey on Drug Use and Health

Introduction

Valid diagnostic criteria sets for any psychiatric disorder, such as alcohol use disorder (AUD), are critical for effectively conducting research on etiology, course, nosology, treatment and prevention. However, available nosological systems for diagnosing AUD have failed to fully derive diagnostic criteria in an empirical and statistical manner (Steinley, Lane, & Sher, 2016; Stevens et al., 2018). For example, AUD diagnosis in *The Diagnostic and Statistical Manual of Mental Disorders* (DSM) has moved from imprecise, ill-defined concepts (e.g., in DSM-I and DSM-II alcoholism; APA, 1952, 1968) to a more highly operationalized approach based on agreement of work group members in later revisions (DSM-III through DSM-5; see also Brown & Barlow, 2009; Kozak & Cuthbert, 2016). For each diagnostic revision, work group members were attentive to the extant empirical literature, commentaries, expert consultation, and field trials in formulating their criteria and diagnostic rules (Cottler & Grant, 2006; Frances & Widiger, 2012; Hasin et al., 2013); however, as with any decision-making process, the ultimate products were the result of a social process influenced, to varying degrees, by political and economic concerns (Frances & Widiger, 2012; Wakefield, 2015). Although this approach is the current standard for developing diagnostic criteria and decision rules for psychiatric diagnoses (e.g., ICD-11; First, Reed, Hyman, & Saxena, 2015), it fails to fully exploit computational methods, which can employ objective, predefined functions for empirically deriving diagnoses at each stage of the decision-making process.

Alternative strategies for developing diagnoses would seem to be strongly justified given critiques of the DSM and similar approaches on the basis of both reliability and validity concerns

(Frances & Nardo, 2013; Phillips et al., 2012; Lilienfeld & Treadway, 2016). For example, considerable concern was raised about criterion redundancy with the combination of alcohol abuse and dependence from DSM-IV to DSM-5. This decision ultimately resulted in a situation whereby an individual can satisfy multiple criteria and reach the diagnostic threshold of two due to a single feature, therefore weakening the validity of the DSM-5 AUD diagnosis (Wakefield, 2015). Further research has demonstrated there are high correlations between AUD criteria such as role impairment and activities given up (tetrachoric $r = 0.85$). This example suggests multiple criteria may reference the same consequence and, as a result, "double-dip" on symptom counts, calling into question the validity of diagnosis and the use of symptom counts to index AUD severity (Martin, Langenbucher, Chung, & Sher, 2014). Concerns such as this point to the potential usefulness of using statistical methods to derive diagnoses.

Diagnostic classification schemes, such as the DSM-5 and *The International Classification of Diseases and Related Health Problems-10* (ICD-10; WHO, 1992) and recently released ICD-11 (First et al., 2015), have tended to prioritize clinical utility over research utility (Kendler et al., 2009; Reed, 2010). Consequently, existing research has failed to adequately exploit empirical techniques and existing databases to derive criteria considered optimal with respect to predicting external correlates. This has resulted in diagnostic criteria sets and rules (DCSRs) that do not necessarily possess a high level of diagnostic efficiency. Diagnostic efficiency can be conceptualized as the relative usefulness or value of specific symptoms for a diagnosis (Pina, Silverman, Alfano, & Saavedra, 2002; Widiger et al., 1984). Given the conceptual overlap in AUD symptoms, demonstrated by authors such as Martin, Langenbucher, Chung, and Sher (2014) as well as Wakefield (2015), it is plausible that not all current AUD criteria are necessary in order to arrive at a diagnosis with equal (or greater) validity. It is also

important to note that the clinical utility and efficiency of diagnoses are dependent upon the setting (e.g., clinical versus non-clinical or outpatient versus emergency room) in which they are used, suggesting that not all criteria may be appropriate for all settings and goals (e.g., Widiger & Trull, 1991; Youngstrom, 2014). The statistical optimization procedure employed by the current paper aims to increase the clinical utility and diagnostic efficiency of AUD diagnosis by reducing the size of the criteria set that needs to be considered for diagnostic purposes while simultaneously optimizing the diagnostic threshold (i.e., number of criteria required for diagnosis) with respect to a known correlate of the disorder. Practically, this could encourage clinicians to learn and use the diagnostic criteria correctly, given research suggesting practitioners are less likely to use the full criteria sets for disorders such as Major Depressive Disorder (e.g., Krupinski & Tiller, 2001) which has a large criteria set size (9 criteria), similar to AUD.

The current diagnostic threshold for DSM-5 AUD is two of eleven criteria. However, some researchers have suggested that this threshold compromises the clinical and research utility of AUD, given it may result in positive diagnoses among those whose substance use has questionable clinical significance and produce marked heterogeneity of symptoms among those a diagnosis (Martin et al., 2011). Additional evidence has suggested that this diagnostic threshold results in overlap in the latent severity, from an Item Response Theory perspective, of AUD between those below the diagnostic threshold (i.e., endorsing less than two criteria) and those at the diagnostic threshold (i.e., endorsing two criteria; Lane & Sher, 2014). Cooper and Balsis (2009) demonstrated that a criteria count of three, which is below the diagnostic threshold for Schizoid Personality Disorder, can have greater latent severity than criteria counts of four, which is at the diagnostic threshold. Therefore, in principle, someone may have a low criterion count,

but a relatively high latent severity of the disorder and vice versa. That is, a person without a diagnosis (based on a criteria count) may actually have more severe psychopathology than an individual with a diagnosis simply based on the criteria they endorse. This is not simply limited to those scores above and below the diagnostic threshold (Cooper & Balsis, 2009; Lane & Sher, 2014). While these issues will likely be relevant for any scale that varies in individual item severity and error variance, as shown in other disorders (e.g., Schizoid Personality Disorder; Cooper & Balsis, 2009), it highlights the importance of where the threshold lies. This implies that diagnostic criteria, specifically in the case of AUD but also in the case of other disorders, are not strictly additive nor interchangeable (as the DSM-5 and other polythetic algorithms assume). In sum, multiple lines of evidence suggest the current AUD criteria set and algorithm leave considerable room for refinement.

To address these challenges, an empirically-based approach, such as the statistical optimization procedure described by Stevens and colleagues (2018; see also Steinley, Lane, & Sher, 2016) and utilized in the current paper, exploits techniques to derive optimal DCSRs through use of complete enumeration of all criteria and thresholds. Empirically-derived criteria sets are considered optimal with respect to predicting relevant external validators, or correlates, robustly and demonstrating generalizability across data sets and populations. This method has the benefit of potentially reducing the number of diagnostic criteria necessary to make a diagnosis of AUD and, thus, increasing diagnostic efficiency and has been used in the context of short-form development (Raffo, Hasin Appelbaum, & Wall, 2018).

The major challenge facing such an approach (i.e., optimization with respect to predicting relevant external criteria) is the question of what criterion or criteria to choose. With respect to AUD there is no single “gold standard,” so it is necessary to choose a criterion that would be

expected to be highly correlated with AUD yet distinct enough to have considerable clinical relevance. A prime candidate for such a criterion is heaviness, or level, of consumption which is separate and distinct from the 11 criteria used to diagnose AUD under the DSM-5 (i.e., drinking more alcohol or over a longer time period than intended; a recurrent desire to cut down on or failed attempts to control use; spending a significant amount of time finding, consuming or recuperating from alcohol; a powerful desire or urge to consume alcohol; failure to uphold important responsibilities at work, school, or home due to use; continued use regardless of social or relational conflicts; activities given up or reduced due to use; persistent use in situations where there is potential for physical harm; sustained use despite awareness of a physical or psychological ailment caused or made worse by alcohol; a need for larger amounts of alcohol to attain the desired effect; and withdrawal). Heaviness of consumption is considered adequate for the purpose of the current study because: (1) heavy use over time is the most parsimonious construct for explaining the neurobiological changes that occur with substance use disorders and for contextualizing the myriad of social and physical consequences that occur in substance users (Rehm & Roerecke, 2013a; Rehm et al., 2013b; Grant et al., 2009) and (2) it shows a strong monotonic relationship with the DSM-5 criteria count and this relationship is more robust than other possible alternative correlates (e.g., general functioning, psychiatric comorbidity; see Dawson, Saha, & Grant, 2010; Lane & Sher, 2014; Saha, Stinson, & Grant, 2007). Additional benefits of this measure are that equivalent consumption measures are typically available in other large, nationally representative surveys, allowing for comparison of derived diagnostic rules, and previous work has demonstrated that similar consumption composites are heritable, influenced by genetic factors that influence heavy drinking, and stable across time (Agrawal, Lynskey, Heath, & Chassin, 2011). In addition, to the extent that a psychological disorder reflects, in part,

distress and impairment (e.g., Wakefield, 1997; Spitzer & Endicott, 1978; APA, 2013), impairment/distress can be viewed as a secondary optimization criterion. Finally, to demonstrate the incremental validity and utility of any new approach, it is necessary to compare its performance to existing standards and relevant alternatives.

The current study sought to exploit the empirical technique described by Stevens and colleagues (2018) to statistically derive an optimal criteria set and diagnostic threshold in the National Survey on Drug Use and Health (NSDUH) data set. The goal of the optimization procedure was to maximize the distance (based on Cohen's *d*) between the mean levels of the optimization criteria (e.g., heaviness of consumption) in those who have an AUD diagnosis (based on a given DCSR) versus those that do not diagnose. The benefit of this methodology over current diagnostic techniques is that it utilizes an objective, data-driven approach to diagnosis where all assumptions and decisions are explicit, transparent, and based on *a priori* considerations. Further, the use of a large, nationally representative data set to achieve this aim has several advantages including: (a) broad generalizability, (b) the assessment of alcohol consumption as well as other relevant correlates (e.g., other co-occurring disorders, treatment usage) for the purposes of external validation, and (c) the different combinations of AUD symptoms (i.e., heterogeneity) occurring in the observed data given the large sample size.

Method

Samples

Data were from the public-use data files of the 2010 and 2013 National Survey on Drug Use and Health (NSDUH), a nationally representative sample of United States civilian, noninstitutionalized individuals aged 12 or older collected by the Substance Abuse and Mental Health Services Administration (SAMHSA, 2011, 2014). Participants were selected using

multistage area probability methods to ensure representativeness. In 2010, 57,313 respondents completed the interview and in 2013, 55,160 separate respondents completed the interview. We limit our sample to those individuals 21 years of age and older resulting in 31,433 and 30,715 respondents for 2010 and 2013, respectively. This restriction was imposed because several epidemiological and clinical studies have demonstrated that AUD diagnostic criteria have limitations when applied to adolescents (e.g., Martin et al., 1996; Martin & Winters, 1998; Winters, Martin, & Chung, 2011). For example, research has demonstrated differential interpretation of diagnostic criteria by adolescents, resulting in false positives and weakened diagnostic validity (Chung & Martin, 2005). Further, participants who did not report consuming at least one drink in the past year were excluded, resulting in total sample sizes of 24,120 and 23,627 for 2010 and 2013, respectively. This exclusion decision is based on the fact that abstainers do not contribute meaningful variance to the outcomes. Full demographic information, including population weighted frequencies, can be found in Table 1. NSDUH's person-level sampling weights were applied to account for NSDUH's independent, multiage area probability sample design (SAMHSA, 2011, 2014). The procedures of this study received approval from the University of XXX Institutional Review Board (Protocol Number: 2003447; Title: Refining the Diagnosis of Alcohol Use Disorder: A Comprehensive Approach).

Measures

Candidate Diagnostic Criteria and Alternate Diagnostic Systems. Alcohol Use Disorder (AUD), according to the DSM-5 (APA, 2013), is characterized by 11 symptoms (APA, 2013). These include: (1) drinking more alcohol or over a longer time period than initially intended ("larger/longer"); (2) a recurrent desire to cut down on alcohol use or failed attempts to control use ("cut down"); (3) spending a significant amount of time finding, consuming or

recuperating from the effects of alcohol ("time spent"); (4) a powerful desire or urge to consume alcohol ("craving"); (5) failure to uphold important responsibilities at work, school, or home due to alcohol use ("failure to fulfill"); (6) continued alcohol use regardless of social or relational conflicts ("social/interpersonal"); (7) important activities given up or reduced due to alcohol use ("give up"); (8) persistent use in situations where there is potential for physical harm to self or others ("hazardous use"); (9) sustained use despite awareness of a physical or psychological ailment caused or made worse by alcohol ("physical/psychological"); (10) a need for larger amounts of alcohol to attain the desired effect ("tolerance"); and (11) withdrawal. To receive a "current" diagnosis of AUD, an individual must have experienced at least two symptoms within the past 12 months. Table 2 shows the endorsement rates for each criterion. Notably, NSDUH fails to assess craving, the criterion added in the DSM-5 revision.

To evaluate our optimal AUD DCSR against other diagnostic rules, we created four additional AUD diagnoses: DSM-$5_{adjusted}$[1], DSM-IV, ICD-10, and Wakefield's harmful dysfunction (Wakefield & Schmitz, 2014, 2015). The prevalence rates of each diagnostic rule are presented in Table 3 and a comparison of the diagnostic rules is presented in Supplemental Table 1.

DSM-$5_{adjusted}$. Given NSDUH excludes craving, we were unable to create the exact DSM-5 diagnosis. Instead, our adjusted DSM-5 diagnostic set includes 10 of the 11 criteria with a diagnostic threshold of two (APA, 2013). Retaining the same threshold guarantees we will not over-diagnose under the DSM-5 framework, but that we will fail to diagnose individuals that the current DSM would have classified if they endorsed craving and one other symptom. A past year DSM-5 AUD diagnosis in those aged 21 years or older who have had at least one alcoholic drink in the past year has an unweighted prevalence rate of 15.89% (N = 3,523) in NESARC-III and an

unweighted prevalence rate of 19.36% (*N* = 4,797) in NESARC Wave 2. A diagnosis of past year DSM-5 AUD without the craving criterion (i.e., DSM-$5_{adjusted}$) results in a prevalence rate of 15.41% (*N* = 3,417) in NESARC-III and 18.00% (*N* = 4,460) in Wave 2. This results in a slightly smaller estimate of AUD when the craving criterion is not included. Despite this, there is a high degree of agreement between a true DSM-5 AUD diagnosis and DSM-$5_{adjusted}$ in NESARC-III (Kappa = 0.96; Tetrachoric = 0.99)[2] and NESARC Wave 2 (Kappa = 0.98; Tetrachoric = 0.99). Table 3 includes prevalence rates for DSM-$5_{adjusted}$ in NSDUH 2010 and 2013 for comparison purposes.

DSM-IV. The DSM-IV (APA, 1994) describes alcohol abuse and dependence as separate, hierarchical diagnoses. That is, if someone is diagnosed with dependence, they cannot also have a diagnosis of abuse, given dependence is devised to be the more severe diagnosis. A diagnosis of alcohol abuse requires endorsement of at least one of the following symptoms: (1) failure to fulfill, (2) hazardous use, (3) social/interpersonal problems, and (4) legal problems. Alcohol dependence requires at least three of the following symptoms: (1) tolerance, (2) quit or cut down, (3) larger/longer, (4) time spent, (5) physical/psychological, (6) give up, and (7) withdrawal. For the purposes of these analyses, we created three variables: DSM-IV abuse, DSM-IV dependence, and DSM-IV AUD (abuse or dependence). Since NSDUH was developed under the DSM-IV framework, complete DSM-IV AUD diagnoses can be made.

ICD-10 Alcohol Dependence. ICD-10 (WHO, 1992) alcohol dependence is described by six symptoms with a diagnostic threshold of three. Some NSDUH AUD items had to be combined to create the diagnosis. These symptoms included: (1) tolerance; (2) cut down OR larger/longer; (3) time spent OR give up; (4) physical/psychological; and (5) withdrawal. A true ICD-10 dependence diagnosis would include craving.

Wakefield Harmful Dysfunction. Wakefield and Schmitz (2014; 2015) developed a diagnostic algorithm assessing harmful dysfunction (HD) AUD which requires both harm (i.e., circumstances judged as negative by sociocultural standards/norms) and dysfunction (i.e., functional impairment) to be present, resulting in lower prevalence rates and improved diagnostic validity (Wakefield & Schmitz, 2015). Wakefield and Schmitz (2015) identified the following symptoms of alcohol dysfunction: withdrawal, drinking to prevent/stop withdrawal, unable to reduce or stop drinking, and craving. Symptoms of alcohol harm include: sacrificing important activities due to alcohol, problems caring for home/family, problems at job/school, health problems related to alcohol consumption, psychological problems, and problems with family or friends. The following symptoms were identified in NSDUH as dysfunction attributed to alcohol use, with a diagnostic threshold of one: wanting/trying to stop or cut down on drinking and withdrawal. Significant distress or impairment from withdrawal symptoms was not assessed in NSDUH and was therefore not included. Further, we were unable to include (1) specific withdrawal symptoms (e.g., shakes or anxiety), (2) relief drinking (i.e., drinking or using a drug/medicine to get over or keep from having withdrawal symptoms), or (3) craving in the dysfunction rule due to the items being unavailable in NSDUH. Harm was present if at least one of the following symptoms was endorsed: important activities given up or reduced, continuing to drink despite physical or psychological problems made worse by alcohol, failure to fulfill important responsibilities, and continuing to drink even though it was causing social/interpersonal problems. Dysfunction and harm were required to meet for Wakefield HD AUD.

Optimization Criteria.

Alcohol Consumption. Individual's self-reported drinking behavior was operationalized with four self-report survey items. These included quantity and frequency of consumption in the past 30 days, frequency of consumption in the past 12 months, and frequency of binge drinking (i.e., five or more drinks on the same occasion where occasion is defined as "at the same time or within a couple hours of each other") in the past 30 days. Frequency of consumption was averaged for past 30 days and past 12 months. The resulting variables were standardized within sex. We then took the mean of the resulting standardized variables to create an alcohol consumption composite. When our composite was created in NESARC-III, the correlation between our composite and that used by Stevens et al., was $r = 0.96$[3], documenting reasonable harmonization of our heaviness of consumption with the corresponding variable used in that study.

Functional Impairment. Functional impairment (i.e., how much emotions, nerves, or mental health have caused difficulties in daily activities) was measured with the World Health Organization Disability Assessment Schedule (WHODAS; Ustun, 2010). NSDUH uses a 13-item reduced version of the scale (Novak, Colpe, Barker, & Gfroerer, 2010; Rehm et al., 1999) that evaluates seven specific difficulties including: remembering important things, concentrating on important things, leaving the house, interpersonal interactions, participation in social activities, taking care of household responsibilities, and maintaining responsibilities at work and school. If participants responded that they did not engage in these activities (where applicable), they received a follow-up question about whether their emotions, nerves, or mental health kept them from engaging in the given activity. Five of the seven difficulties had "did not engage in this activity" as an option. These items were not included in the mean score if participants failed to engage in three or more of the five applicable activities. In such a case, the participant's total

WHODAS score was coded as missing (2010: *n* = 173, 0.72%; 2013: *n* = 165, 0.70%). For individuals who reported they did not engage in only one or two of the five applicable activities, their total scores were imputed by estimating the mean of the available cases (Enders, 2010). The resulting WHODAS mean score was then standardized. Alphas for the 2010 and 2013 WHODAS were 0.90 and 0.91, respectively.

External validators. The importance of external validation in this context is to demonstrate that our newly constructed optimal diagnoses are associated with a range of clinically and/or theoretically relevant variables, including those of the ostensibly same construct (Grimm & Widaman, 2012). An additional goal of the external validation was to evaluate whether there is incremental validity of the optimal solution above and beyond the traditional diagnostic rules (e.g., DSM-IV, ICD-10) with respect to the prediction of these clinically and/or theoretically relevant external criteria.

The purpose of the external validation procedure in the current study was to determine how various diagnostic approaches (i.e., alternative existing diagnostic rule, our derived optimal solutions) performed with respect to each other in predicting relevant criteria. We were able to identify a range of external validators (i.e., known correlates of AUD) including: past year treatment usage including formal treatment (e.g., rehabilitation facility, visiting a counselor), and informal treatment (e.g., Alcoholics Anonymous); psychopathology that commonly co-occurs with AUD (e.g., past year mood disorder and drug use disorder); and early onset of drinking.

Optimization Procedure

The statistical optimization approach used in the current study were developed by Stevens and colleagues (2018) and represents a systematic extension of our earlier approach described by Steinley et al. (2016). The approach proceeds as follows:

Step 1. As suggested by Rodríguez and colleagues (2010), each of the datasets (NSDUH 2010 and 2013) were randomly divided into five non-overlapping folds, resulting 10 total folds with ~4,000 observations per fold. Folds are subsets of the full datasets that the optimization procedure can be performed on independently, allowing us to examine how each DCSR performs on partitions of the data. Splitting the data in such a way is a cross-validation method used in optimization procedures to safeguard against over fitting to the full dataset (Stone, 1974).

Step 2. We specified, *a priori*, the optimization criteria and weighting, item sets, and base rate. Our optimization criteria, as described previously, included alcohol consumption and functional impairment. The weighting ranged from 100% consumption, 0% functional impairment to 0% consumption, 100% functional impairment in 25% intervals (e.g., consumption-impairment: 100-0, 75-25, 50-50, 25-75, 0-100). Different patterns of weighting allowed us to determine the amount each optimization criterion contributes to the diagnosis of AUD and the consistency of solutions across varying specifications. Our item set included the DSM-5 AUD symptoms described above except for craving, resulting in 10 items. The optimization procedure therefore assessed every combination of symptoms among the 10 items across all diagnostic thresholds. Since heaviness of consumption is highly correlated with the symptom count, a minimum base rate of the disorder must be designated. This step ensures that the selected optimal diagnostic set is not restricted to the very extreme end of AUD. For these reasons, the AUD minimum base rate was set at 10%, consistent with the weighted DSM-IV AUD (abuse or dependence) prevalence rates in NSDUH 2010 (9.9%) and 2013 (9.5%) for past year drinkers aged 21 years or older. This base rate ensures that in at least one fold of the data, a minimum of 10% of the observations diagnose with the given diagnostic rule[4]. This allowed us to understand how the diagnosis performs at this proportion or above within the distribution.

DSM-IV was chosen over DSM-5 to inform the selected base rate given NSDUH does not include the full DSM-5 criteria set.

Step 3. At this step, Cohen's *d* was calculated for every diagnostic item set combination varying the number of symptoms and diagnostic thresholds (both ranging from one to 10). This results in a total of $2^{10} - 1 = 1{,}023$ different combinations, not including the empty set, to be examined at a diagnostic threshold of one. Supplemental Table 2 summarizes the number of combinations within each set size and threshold. When the diagnostic threshold is varied, this results in a total of 5,120 possible combinations and 5,120 estimates of Cohen's *d*, allowing us to assess the separation of diagnostic classifications for all DCSRs.

Step 4. The resulting Cohen's *d* for every DCSR, within each fold, were then ranked. High values of Cohen's *d* (i.e., the diagnostic groups had greater separation) were assigned lower numerical rankings and vice versa, allowing for more robust solutions when comparing across the folds. For example, the largest Cohen's *d* within each fold received a rank of 1 and the smallest received a rank of 5,120. DCSRs were considered eligible as the overall optimal solution if the prevalence rate was greater than or equal to the pre-specified base rate of 10% in at least one fold.

Step 5. Following the ranking procedure within each fold, these rankings were averaged across all folds. The DCSR with the lowest average ranking across the folds was identified as the overall *best* solution. That is, the overall optimal solution is the average lowest ranking solution across the 10 folds. Since a possible goal of using this optimization procedure is to decrease the symptom set size, secondary optimal solutions were selected as diagnostic rule with at most six symptoms.

Step 6. The overall optimal solution and secondary optimal solution produced by the procedure were then compared to the other existing AUD diagnostic rules and techniques described above (e.g., ICD, Wakefield) using the Adjusted Rand Index (ARI; Hubert & Arabie, 1985). The ARI is a measure of similarity that is adjusted for chance grouping whereby higher values represent greater similarity. Although there is no rule of thumb for ARI, values above 0.6 are generally considered adequate.

Further, performance of the overall optimal solution, compared to existing AUD diagnostic rules, was examined via the use of external validators. The external validation was conducted in two steps using SAS's PROC SURVEYLOGISTIC. The first set of analyses evaluated the ability of each DCSR to predict each relevant outcome alone. Odds ratios and confidence intervals were calculated to determine if the diagnostic rule was a significant predictor of the external validator. The second set of analyses examined the degree of incremental validity of the optimal solutions over alternative diagnostic rules. Degree of incremental validity was determined by the change in the c statistic which is a measure equivalent to the receiver operating characteristic curve (ROC). The c statistic ranges from 0.5 to 1, where 0.5 corresponds to the model randomly predicting the response, and 1 corresponds to the model perfectly discriminating the response. The fit of the more parsimonious model (the model with the alternative diagnostic rule only) compared to the full model (the model with the alternative diagnostic rule AND the new optimal solution) can be compared using a Chi-square difference test. A significant Chi-square estimate indicates that the full model is significantly more predictive than the reduced model.

Results

The overall NSDUH 2010/2013 optimal solution is described in Tables 4 and 5. Although results in Table 4 are presented for the range of optimization criteria weights, we propose that the 100% consumption, 0% functional impairment results should be considered first since identical optimal solutions were identified at weighting schemes ranging from 50% consumption, 50% functional impairment to 100% consumption, 0% functional impairment. Using the more parsimonious method would imply that functional impairment is adding little information to the diagnosis with the given symptoms. Therefore, the overall optimal solution includes the following nine criteria, with a diagnostic threshold of three: larger/longer, cut down, time spent, failure to fulfill, social/interpersonal, hazardous use, physical/psychological, tolerance, and withdrawal. When considering secondary optimal solutions, the second best overall solution includes the following five criteria with a diagnostic threshold of two: cut down, time spent, social/interpersonal, hazardous use, and withdrawal (Table 5). Both optimal solutions result in a smaller criteria set size when compared to DSM-5, albeit minimally so in the case of the overall optimal solution.

Table 5 provides a summary of the top five performing DCSRs with respective average Cohen's *d*s. While the use of cut-offs for Cohen's *d* is not ideal, cut-offs can aid in the interpretation of the statistic. Cohen's *d* is usually quantified as small (0.2), medium (0.5) or large (0.8) (Cohen, 1977). This table demonstrates that the secondary optimal solution provides a more parsimonious solution than the overall optimal solution while still ranking similarly in terms of Cohen's *d*. The ARI, indicating agreement among bivariate solutions, for the overall and second overall solution was 0.77 while the weighted Phi coefficient, indicating degree of association, was 0.80. These, along with Kappa coefficients and tetrachoric correlations, are also provided for the top two optimal solutions across the various diagnostic rules (Tables 6 and 7).

The second (and most parsimonious) overall optimal solution, does not result in a loss of significant information, and demonstrates excellent agreement and a very strong association with the overall optimal solution. To further determine whether the more parsimonious solution could be favored against the other two solutions, we compared each to a range of external validators.

External validation

SAS's PROC SURVEYLOGISTIC was used to estimate the weighted odds ratios (ORs) and degree of incremental validity (change in c [Δc]) for each pre-established diagnostic rule and new optimal solutions across the range of external validators in NSDUH 2013, controlling for sex and age (Table 8). In addition, chi-square difference tests were conducted to determine whether there was a significant change in model fit when the overall and second overall optimal solutions were added to a model including one of the preexisting diagnostic rules (e.g., DSM-IV).

Results demonstrated that the first and second overall optimal solutions perform similarly in their prediction of the range of external validators, as indicated by comparable odds ratios, when compared to preexisting diagnostic rules. Although our recommendation is that the overall optimal solution be used, based on its overall Cohen's *d*, the second overall optimal solution appears to perform similarly with the added benefit of considerably greater parsimony (i.e., 2 out of 5 criteria) and diagnostic efficiency. When the first and second overall optimal solutions are evaluated for degree of incremental validity and improvement in model fit (via a Chi-Square difference test), the two optimal solutions almost always result in a positive Δc and a significant chi-square difference. While it may initially seem unexpected that the ORs are comparable across the different DCSRs yet the Chi-Square difference tests are significant, this is not surprising given the extremely large sample size and the sensitivity of the Chi-Square statistic

(Fornell & Larcker, 1981). Given these optimal solutions performed as well as preexisting AUD diagnostic rules in terms of predicting relevant outcomes, it is important to note that the optimal solutions provided the added benefit of a smaller set size and therefore increased diagnostic efficiency.

Discussion

The overall optimal solution includes the following nine criteria with a diagnostic threshold of three: (1) drinking larger amounts or over a longer period than intended; (2) attempts to quit or cut down cut down; (3) a significant amount of time spent drinking, being sick, or getting over the effects of alcohol; (4) failure to fulfil important activities; (5) social or interpersonal problems resulting from drinking; (6) hazardous use; (7) physical/psychological problems; (8) tolerance; and (9) withdrawal. This excludes the symptom of giving up or reducing important activities due to alcohol use. The second overall solution consisted of a subset of five criteria (i.e., [1] cut down; [2] time spent; [3] social/interpersonal; [4] hazardous use; and [5] withdrawal), with a diagnostic threshold of two. This second overall solution performs comparably across a range of external validators.

While the preference would be to favor the overall optimal solution from a statistical point of view, the second overall optimal solution and its smaller set size have important clinical implications. Clearly, there is a smaller cognitive load (i.e., degree of mental activity imposed on working memory; Burgess, 2010) on the part of the clinician, less response burden (i.e., degree of effort required to answer a questionnaire or an interview; Rolstad, Adler, & Ryden, 2011) for the patient, and increased diagnostic efficiency with a criterion set of 5 as compared to 9 (or more). Issues such as cognitive load and response burden have been noted as important factors, particularly in the primary care literature. For example, in several experiments, high cognitive

load has been demonstrated to impact diagnostic decision making (e.g., Cumming & Harris, 2001), problem solving (e.g., Sweller, 1988), and learning (Weidman & Baker, 2015). In general, working memory can handle only a very limited number of elements – possibly no more than three to five (e.g., Cowan, 2010). This suggests that the larger number of criteria an assessor has to consider at a given time, the larger the cognitive demand and, as a result, the more likely diagnostic errors are to result (e.g., Thammasitboon & Cutrer, 2013). These findings also imply that the minimization of unnecessary or redundant information can decrease cognitive load (e.g., van Merrienboer & Sweller, 2009), particularly in the case of reduced set sizes, leading to improved information processing and subsequent storage into long-term memory (Schnotz & Kurschner, 2007). As a result, given its comparable statistical performance to the overall optimal solution, the second overall optimal solution may be preferred depending on factors such as the goals of the diagnosis and the setting (e.g., hospital emergency room versus a comprehensive diagnostic clinic).

These results stand in comparison to those found by Stevens and colleagues (2018) which demonstrated, in NESARC Wave 2 (Grant & Kaplan, 2005) and III (Grant et al., 2014), that seven criteria (tolerance, larger/longer, time spent, give up, physical/psychological, craving, and social/interpersonal) with a diagnostic threshold of three was optimal. It is important to consider how the inclusion of craving resulted in a different pattern of results. Craving has been found to be an important symptom of AUD (MacKillop et al., 2010) and the lack of assessment in NSDUH could result in different optimal sets than would have been determined with its inclusion[5]. We view this as a benefit rather than a limitation given it allows an examination of how the statistical optimization procedure performs in a comparable, nationally representative sample with slightly varying AUD criteria. Further, while NESARC certainly has the benefit of

including craving, these are other meaningful differences between the two data sets that make this a worthwhile endeavor. NESARC has been criticized for methodological issues such as inflated incident rates (e.g., Caetano, 2015) and problems with the assessment of the withdrawal criterion (e.g., Boness, Lane, & Sher, 2016), pointing to the importance of examining how the optimal solution differs across data sets to determine how robust the optimal solutions are across different data sets and assessment instruments. The results of the current paper, therefore, provide an important comparison and extension of the results described by Stevens and colleagues (2018).

The overall optimal solution, with a criterion set size of nine and a diagnostic threshold of three, addresses some of the issues related to the DSM-5's two of eleven rule. Martin and colleagues (2011) have argued that the two of eleven rule results in (1) diagnosed cases with a lack of clinical significance (i.e., very mild cases of AUD) and (2) creates significant heterogeneity. Our optimal solution potentially addresses some of these issues by raising the diagnostic threshold to three and reducing the set size to nine. A higher diagnostic threshold and lower criteria set size may capture cases of AUD that are more clinically meaningful while also reducing the heterogeneity among diagnosed cases.

As with any modifications to criteria set sizes and diagnostic thresholds, it is important to consider the potential for diagnostic orphans (e.g., Hasin & Paykin, 1998; Martin, Chung, & Langenbucher, 2008) and impostors (e.g., Langenbucher, Martin, Hasin, & Helzer, 1996; Martin, Chung, & Langenbucher, 2008) which are traditionally described in the context of DSM-IV dependence though are applicable to DSM-5 AUD as well. Diagnostic orphans describe individuals with sub-threshold AUD symptoms. Diagnostic impostors describe individuals who diagnose with AUD despite low levels of alcohol use. In general, we feel it is more important to

guard against the more prevalent diagnostic imposters by requiring a diagnostic threshold of at least two in our optimal solutions. To the extent that any diagnostic system will have sensitivity or specificity less than one, there will always be a degree of misdiagnosis. As such, we must await a true gold standard to fully resolve this issue. An additional benefit of the optimization procedure is the fact that it optimizes on consumption and therefore is less likely to be plagued by the issues related to misdiagnosis within the current DSM-5 framework (e.g., problems with the withdrawal criterion; further discussed below). Consumption also has the advantage of being associated with neurobiological changes as well as social and physical consequences that occur in substance users (e.g., Rehm & Roerecke, 2013a), as described previously. Further, as previously mentioned, similar consumption composites are heritable, influenced by genetic factors that influence heavy drinking, and stable across time (Agrawal, Lynskey, Heath, & Chassin, 2011), suggesting that the use of consumption as the primary optimization criterion may be more objective and valid than the diagnostic criteria themselves.

It should be noted that the base rate of 10% was selected because of its association with the diagnostic rates in DSM-IV AUD, however, any base rate could potentially be examined. Lower declared base rates will result in higher levels of Cohen's *d* simply because the degree of separation between diagnostic groups will increase as more severe populations (lower base rates) are considered. For example, if the base rate was set to coincide with the ICD-10 prevalence within NSDUH (about 5%), we would classify only the most severe individuals with AUD that have consumption rates within the top 5% of the sample. Additionally, it is important to note that the optimization procedure does not account for population weighting used in NSDUH. While we do not believe this is a particularly important limitation, there may be some degree of imprecision when generalizing as a result.

One limitation of the current study is the fact that the NSDUH data sets do not include a wide range of AUD candidate criteria (e.g., the exclusion of craving). By limiting ourselves to the criteria found in common diagnostic surveys, the universe of potential criteria that could be evaluated is limited. This had two important effects: (1) it prevented us from creating a true DSM-5 diagnosis for comparison purposes, and (2) it restricted our ability to derive criteria sets beyond the 10 criteria assessed in NSDUH. The impact of this limitation is apparent in light of the optimal solution found by Stevens and colleagues in NESARC Wave 2 and NESARC-III, which included the craving criterion and ultimately resulted in an optimal criterion set with fewer items. This demonstrates the importance of the population of candidate items under consideration and how the addition of particular criteria, such as craving, can impact the overall optional solution. While we recognize this issue, we also acknowledge that the NSDUH dataset is ideal for optimization due to the fact that it is a nationally representative survey and its annual fielding allows us to examine consistency of findings across a large number of random subsamples (i.e., folds).

A further caveat is found in the withdrawal criterion. DSM-5 defines withdrawal as (1) cessation of or reduction in heavy or prolonged alcohol use plus (2) the presence of two or more withdrawal symptoms (e.g., autonomic hyperactivity, increased hand tremor, sleep trouble) and (3) significant impairment of distress OR the use of alcohol to relieve or avoid withdrawal symptoms (APA, 2013). However, NSDUH does not assess withdrawal symptoms separately and instead asks: "…did you have 2 or more of these symptoms after you cut back or stopped drinking alcohol?" More critically, participants were not assessed for withdrawal if they responded "no" to the item: "In the past 12 months, did you cut down or stop drinking at least one time?" This impacts the results as those who did not try to quit or cut down were excluded

from the optimization analyses due to missing data. Consequently, withdrawal is likely more associated with the symptom "cut down" than it would have been otherwise. In addition, drinking to avoid or relieve withdrawal symptoms was not assessed by NSDUH. NSDUH's partial assessment of withdrawal may jeopardize the validity of this criterion. However, other nationally representative surveys, such as NESARC, are also plagued by issues with their assessment of withdrawal (e.g., Boness, Lane, & Sher, 2016; Karriker-Jaffee, Witbrodt, & Greenfield, 2015). Thus, although there may appear to be some divergence between the definition of withdrawal in the DSM and its operationalization in NSDUH, the wording in NSDUH could potentially reduce the problems of false positive withdrawal symptoms common in other assessments such as NESARC.

We also note that one of the criteria that consistently appeared in the optimal solutions was hazardous use. This may be cause for concern given the known psychometric and conceptual issues with this criterion (see Martin, Sher, & Chung, 2011). For example, hazardous use demonstrates only modest discrimination in Item Response Theory analyses (Lynskey & Agrawal, 2007) and it is defined solely by risky behavior (e.g., driving under the influence) rather than the key components of substance use disorders such as substance-related consequences, physiological features, or compulsive use. Thus, although hazardous use performed well statistically, its appearance in our optimal solutions highlights differences between rational and "brute force" statistical approaches.

The issues with withdrawal and hazardous use demonstrate an important consideration of the statistical optimization procedure—the procedure does not guard against conceptually problematic diagnostic items. Therefore, the procedure cannot guarantee the optimal solution, *by itself,* has construct validity as it is based solely on a set of preexisting items, such as those

included in the NSDUH data set. This points to the importance of using construct valid criterion sets. Optimization procedures such as the one utilized in the current paper are only as good as their candidate criteria, pointing to the importance of developing construct valid AUD criteria.

Though we are sensitive to the particulars of each data set employed (see Lane et al., 2016) and feel that more definitive analyses await more extensive validation across instruments and larger item pools for evaluating alcohol-related pathology more extensively, we believe we can derive DCSRs that are more efficient than those of existing standards without loss of validity. Perhaps, most critically, this approach is not subject to the influence of the political and subjective processes that can influence standard practices for developing criteria sets and algorithms. Even when such a process is employed (e.g., via the DSM or ICD approach), analyses such as those presented here can provide valuable information to developers of rationally designed approaches by providing critical information on diagnostic efficiency. The optimization approach also has the advantage of selecting a robust solution given it utilizes cross-validation techniques that protect against overfitting.

Future Directions

The current paper and optimization procedure described offer a viable option for the statistical optimization of AUD criteria. We believe this methodology is compatible with movements such as the Research Domain Criteria (RDoC; e.g., Insel et al., 2010) and Alcohol Addiction Research Domain Criteria (AARDoC; Litten et al., 2015). These approaches aim to characterize clinical disorders by focusing on mechanisms rather than clinical description or symptoms. As these frameworks become more developed and relevant transdiagnostic dimensions are identified, the optimization approach described in the current paper will offer a promising technique for determining construct valid optimal DCSRs. Given these frameworks

are in their early development and various methodological and conceptual challenges remain (Miller & Rockstroh, 2013; Sher, 2015), the use of extant diagnostic frameworks in the current paper only provide an important starting place. An additional future consideration might involve the evaluation of complex combinations or decision rules (e.g., 1 of the following 3 symptoms, and at least 2 of the following 4 symptoms), especially given AUD criteria are not additive nor interchangeable. However, this will significantly increase the number of enumerations and therefore require more computational power.

Although the current paper used alcohol consumption and functional impairment as optimization criteria, future work may benefit from using a measure of chronicity (i.e. persistence of AUD diagnosis) such as that used by Steinley and colleagues (2016). Chronicity has been recognized as an important component of diagnosis since the time of Kraepelin (1899) and may be a more valid criterion over alcohol consumption and functional impairment. Relatedly, while the use of a nationally representative sample is ideal, the cross-sectional nature of this data in the statistical optimization procedure necessarily limits the conclusions that can be made regarding how well the current consumption measure describes the typical drinking of participants. It is possible that the criteria that are most predictive of long-term harm are not the same as the criteria that are associated with current consumption/functional impairment. Without highly resolved prospective measures of all constructs, it is not possible to adequately characterize the time-bound functional relations among these constructs. In addition, future studies might consider how the optimization procedure may perform differently across groups. For example, it may be the case that optimal solution differs based on age, sex, or ethnicity.

In conclusion, the statistical optimization procedure described offers a viable option for improving the efficiency of the AUD diagnosis. However, it is important to keep in mind the

considerations described above. In the future, it is imperative to focus on improving the construct validity of AUD diagnostic criteria, potentially by embracing new frameworks such as RDoC. As the field continues to improve the construct validity of AUD diagnoses, optimization approaches will become even more useful to researchers.

Footnotes

[1] DSM-$5_{adjusted}$ does not include craving and therefore requires two of the ten (i.e., larger/longer, cut down, time spent, failure to fulfil, social/interpersonal, give up, hazardous use, physical/psychological, tolerance; and withdrawal) AUD criteria are endorsed.

[2] To demonstrate the agreement more clearly, in NESARC-III, only 7% fewer participants are diagnosed under DSM-$5_{adjusted}$ compared to a true DSM-5 diagnosis.

[3] The consumption composite in NESARC Wave 2 and NESARC-III was created using past year items measuring: 1) frequency of consumption, 2) frequency of drinking the maximum number of drinks consumed, 3) quantity consumed, 4) largest number of drinks consumed in a single day, 5) frequency of intoxication, 6) frequency of binge drinking (as defined as 4+ drinks for females and 5+ drinks for males over a two hour period; NIAAA, n.d.), and 7) frequency of exceeding daily limits (>3 drinks in a single day for females, >4 drinks in a single day for males; NIAAA, n.d.). The resulting composite was standardized by sex.

[4] Declaring a base rate in at least one fold is necessary to ensure that the program does not select an optimal solution only separating those endorsing all symptoms versus those endorsing no symptoms. A solution cannot be selected as the optimal solution if it falls below the defined base rate in all folds. Requiring this constraint within at least one fold allows for variation below the declared base rate. This is important because we do not view the input base rate as a hard cutoff for identifying an optimal solution, but rather a rough estimate of the diagnostic rate within the population informed by prior research.

[5] The importance of this is demonstrated by the fact that 88% of the top 100 diagnostic rules produced by the optimization procedure in NESARC-III included craving.

References

Agrawal, A., Lynskey, M. T., Heath, A. C., & Chassin, L. (2011). Developing a genetically informative measure of alcohol consumption using past-12-month indices. *Journal of Studies on Alcohol and Drugs, 72*(3), 444-452.American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th Ed.).* Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Arlington: VA: American Psychiatric Publishing.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th Ed.).* Washington, DC: Author.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders (2nd Ed.).* Washington, DC: American Psychiatric Association.

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders (1st Ed.).* Washington, DC: American Psychiatric Association.

Boness, C. L., Lane, S. P., & Sher, K. J. (2016). Assessment of withdrawal and hangover is confounded in the Alcohol Use Disorder and Associated Disabilities Interview Schedule: Withdrawal prevalence is likely inflated. *Alcoholism: Clinical and Experimental Research, 5*, 1–9.

Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment*, *21*(3), 256–71.

Burgess, D. J. (2010). Are providers more likely to contribute to healthcare disparities under high levels of cognitive load? How features of the healthcare setting may lead to biases in medical decision making. *Medical Decision Making, 30*(2), 246-257.

Caetano, R. (2015). A decade after NESARC: What has it told us? *Addiction, 110*(3), 375-377.

Chung, T., & Martin, C. S. (2005). What were they thinking? Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug and Alcohol Dependence, 80*(2), 191–200.

Cohen, J. (1977). *Statistical power analysis for behavioral sciences (revised Ed.)*. New York: Academic Press.

Cooper, L. D., & Balsis, S. (2009). When less is more: How fewer diagnostic criteria can indicate greater severity. *Psychological Assessment, 21*(3), 285.

Cottler, L. B., & Grant, B. F. (2006). Characteristics of nosologically informative data sets that address key diagnostic issues facing the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) and International Classification of Diseases, eleventh edition (ICD-11) substance use disorders workgroups. *Addiction, 101*, 161–169.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current Directions in Psychological Science, 19*(1), 51-57.

Cumming, S. R., & Harris, L. M. (2001). The impact of anxiety on the accuracy of diagnostic decision-making. *Stress and Health, 17*(5), 281-286.

Dawson, D. A., Saha, T. D., Grant, B. F. (2010). A multidimensional assessment of the validity and utility of alcohol use disorder severity as determined by item response theory models. *Drug and Alcohol Dependence, 107*, 31-38.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.

First, M. B., Reed, G. M., Hyman, S. E., & Saxena, S. (2015). The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders. *World Psychiatry, 14*(1), 82-90.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 39-50.

Frances, A. J., & Nardo, J. M. (2013). ICD-11 should not repeat the mistakes made by DSM-5. *British Journal of Psychiatry*, *203*(1), 1–2.

Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology, 8*, 109–130.

Grant, J. D., Agrawal, A., Bucholz, K. K., Madden, P. A., Pergadia, M. L., Nelson, E. C., ... & Whitfield, J. B. (2009). Alcohol consumption indices of genetic risk for alcohol dependence. *Biological Psychiatry*, *66*(8), 795-800.

Grant, B. F., Amsbary, M., Chu, A., Sigman, R., Kali, J., Sugawana, Y., ... & Chou, P. S. (2014). Source and accuracy statement: National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III). *Rockville, MD: National Institute on Alcohol Abuse and Alcoholism*.

Grant, B.F., & Kaplan, K.D. (2005). Source and Accuracy Statement for the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *National Institute on Alcohol Abuse and Alcoholism, Rockville, MD.*

Grimm, K. J., & Widaman, K. F. (2012). Construct validity. In *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 621– 642). Washington, DC: American Psychological Association.

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Compton, W.M., Crowley, T., Ling, W., Petry, N. M., Schuckit, M., & Grant, B. F. (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry*, *170*(8), 834-851.

Hasin, D., & Paykin, A. (1998). Dependence symptoms but no diagnosis: diagnostic orphans' in a community sample. *Drug and Alcohol Dependence, 50*(1), 19-26.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193-218.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry, 167*(7), 748-751.

Karriker-Jaffe, K. J., Witbrodt, J., & Greenfield, T. K. (2015). Refining measures of alcohol problems for general population surveys. *Alcoholism: Clinical and Experimental Research*, *39*(2), 363-370.

Kendler, K. S., Aggen, S. H., Prescott, C, Crabbe, J., & Neale, M. C. (2011). Evidence for multiple genetic factors underlying the DSM-IV criteria for alcohol dependence. *Molecular Psychiatry*, *17*(12), 1306–1315.

Kraepelin, E. (1899) *Psychiatrie.* Leipzig: Barth

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH Research Domain Criteria initiative: Background, issues, and pragmatics. *Psychophysiology*, *53*(3), 286–297.

Krupinski, J., & Tiller, J. (2001). The identification and treatment of depression by general practitioners. *Australian & New Zealand Journal of Psychiatry, 53,* 286-297.

Lane, S. P., & Sher, K. J. (2014). Limits of current approaches to diagnosis severity based on criterion counts: An example with DSM-5 alcohol use disorder. *Clinical Psychological Science*, *3*(6), 819-835.

Lane, S. P., Steinley, D., & Sher, K. J. (2016). Meta-analysis of DSM alcohol use disorder criteria severities: Structural consistency is only 'skin deep.' *Psychological medicine*, *46*(8), 1769-1784.

Langenbucher, J. W., Martin, C. S., Hasin, D. S., & Helzer, J. E. (1996). Alcohol abuse: Adding content to category. *Alcoholism: Clinical and Experimental Research, 20*, 270A–275A.

Lilienfeld, S. O., & Treadway, M. T. (2016). Clashing diagnostic approaches: DSM-ICD versus RDoC. *Annual Review of Clinical Psychology*, *12*(1), 435–463.

Litten, R. Z., Ryan, M. L., Falk, D. E., Reilly, M., Fertig, J. B., & Koob, G. F. (2015). Heterogeneity of alcohol use disorder: Understanding mechanisms to advance personalized treatment. *Alcoholism: Clinical and Experimental Research*, *39*(4), 579–584.

Lynskey, M. T., & Agrawal, A. (2007). Psychometric properties of DSM assessments of illicit drug abuse and dependence: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Psychological Medicine*, *37*(9), 1345-1355.

MacKillop, J., Miranda Jr, R., Monti, P. M., Ray, L. A., Murphy, J. G., ... & Gwaltney, C. J. (2010). Alcohol demand, delayed reward discounting, and craving in relation to drinking and alcohol use disorders. *Journal of Abnormal Psychology*, *119*(1), 106.

Martin, C. S., Chung, T., & Langenbucher, J. W. (2008). How should we revise diagnostic criteria for substance use disorders in the DSM-V?. *Journal of Abnormal Psychology, 117*(3), 561.

Martin, C. S., Langenbucher, J. W., Chung, T., & Sher, K. J. (2014). Truth or consequences in the diagnosis of substance use disorders. *Addiction, 109*(11), 1773-1778.

Martin, C. S., Langenbucher, J. W., Kaczynski, N. A., & Chung, T. (1996). Staging in the onset of DSM-IV alcohol symptoms in adolescents: survival/hazard analyses. *Journal of Studies on Alcohol, 57*(5), 549-558.

Martin, C. S., Sher, K. J., & Chung, T. (2011). Hazardous use should not be a diagnostic criterion for substance use disorders in DSM-5. *Journal of Studies on Alcohol and Drugs*, *72*(4), 685-686.

Martin, C. S., Steinley, D. L., VergÈs, A., & Sher, K. J. (2011). Letter to the editor: The proposed 2/11 symptom algorithm for DSM-5 substance-use disorders is too lenient. *Psychological Medicine, 41*(9), 2008-2010.

Martin, C. S., & Winters, K. C. (1998). Diagnosis and assessment of alcohol use disorders among adolescents. *Alcohol Health and Research World, 22*, 95–105.

Miller, G. A, & Rockstroh, B. (2013). Endophenotypes in psychopathology research: where do we stand? *Annual Review of Clinical Psychology, 9*(December), 177–213

National Institute on Alcohol Abuse and Alcoholism [NIAAA]. (n.d.). Drinking levels defined. Retrieved from https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research, 19*(S1), 49-60.

Phillips, J., Frances, A., Cerullo, M. A., Chardavoyne, J., Decker, H. S., First, M. B., … Zachar, P. (2012). The six most essential questions in psychiatric diagnosis: A pluralogue part 2:

Issues of conservatism and pragmatism in psychiatric diagnosis. *Philosophy, Ethics, and Humanities in Medicine, 7*(1), 8.

Pina, A. A., Silverman, W. K., Alfano, C. A., & Saavedra, L. M. (2002). Diagnostic efficiency of symptoms in the diagnosis of DSM- IV: Generalized anxiety disorder in youth. *Journal of Child Psychology and Psychiatry*, *43*(7), 959-967.

Raffo, C. D., Hasin, D. S., Appelbaum, P., & Wall, M. M. (2018). A data-driven method for identifying shorter symptom criteria sets: The case for DSM-5 alcohol use disorder. *Psychological Medicine,* 1-9.

Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's International Classification of mental disorders. *Professional Psychology: Research and Practice*, *41*(6), 457.

Rehm, J. & Roerecke, M. (2013a). Reduction of drinking in problem drinkers and all-cause mortality. *Alcohol and Alcoholism, 48*, 509-513.

Rehm, J., Marmet, S., Anderson, P., Gual, A., Kraus, L., Nutt, D. J., Room, R., Samokhvalov, A.V., ... Wiers, R.W., & Gmel, G. (2013b). Defining substance use disorders: Do we really need more than heavy use? *Alcohol and Alcoholism, 48*, 633-640.

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. D. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research*, *8*(2), 110-122.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569-575.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101-1108.

Saha, T. D., Stinson, F. S., & Grant, B. F. (2007). The role of alcohol consumption in future classifications of alcohol use disorders. *Drug and Alcohol Dependence, 89*, 82–92.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Sher, K. J. (2015). Moving the Alcohol Addiction RDoC forward. *Alcoholism: Clinical and Experimental Research*, *39*(4), 579–584.

Spitzer, R. L., & Endicott, J. (1978). Medical and mental disorder: Proposed definition and criteria. *Critical Issues in Psychiatric Diagnosis*, 15-39.

Steinley, D., Lane, S. P., & Sher, K. J. (2016). Determining optimal diagnostic criteria through chronicity and comorbidity. *In Silico Pharmacology, 4*(1), 1-12.

Stevens, J.E., Steinley, D., Boness, C.L. Trull, T., Wood, P., & Sher, K.J. (2018). Combinatorial optimization of classification decisions: An application to refine psychiatric diagnoses. https://doi.org/10.31234/OSF.IO/JPNMF

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 111-147.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2011). Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings, NSDUH Series H-41, HHS Publication No.(SMA) 11-4658. *Rockville, MD: Substance Abuse and Mental Health Services Administration, 201*.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2014). Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings,

NSDUH Series H-48, HHS Publication No.(SMA) 14–4863. *Rockville, MD: Substance Abuse and Mental Health Services Administration*, 13-4795.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Thammasitboon, S., & Cutrer, W. B. (2013). Diagnostic decision-making and strategies to improve diagnosis. *Current Problems in Pediatric and Adolescent Health Care*, *43*(9), 232-241.

Üstün, T. B. (Ed.). (2010). *Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0*. World Health Organization.

Wakefield, J. C. (1997). Diagnosing DSM-IV—Part I: DSM-IV and the concept of disorder. *Behaviour Research and Therapy, 35*(7), 633-649.

Wakefield, J. C. (2015). DSM-5 substance use disorder: How conceptual missteps weakened the foundations of the addictive disorders field. *Acta Psychiatrica Scandinavica, 132*(5), 327–334.

Wakefield, J. C., & Schmitz, M. F. (2014). How many people have alcohol use disorders? Using the harmful dysfunction analysis to reconcile prevalence estimates in two community surveys. *Frontiers in Psychiatry, 5*, 10.

Wakefield, J. C., & Schmitz, M. F. (2015). The harmful dysfunction model of alcohol use disorder: Revised criteria to improve the validity of diagnosis and prevalence estimates. *Addiction, 110*(6), 931–942.

Weidman, J., & Baker, K. (2015). The cognitive science of learning: Concepts and strategies for the educator and learner. *Anesthesia & Analgesia*, *121*(6), 1586-1599.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and DSM-III. *Archives of General Psychiatry, 41*(10), 1005-1012.

Widiger, T. A., & Trull, T. J. (1991). Diagnosis and clinical assessment. *Annual Review of Psychology, 42*(1), 109-133.

Winters, K. C., Martin, C. S., & Chung, T. (2011). Substance use disorders in DSM-V when applied to adolescents. *Addiction, 106*(5), 882-884.

World Health Organization [WHO]. (1992). *The ICD-10 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines*. Geneva: WHO.

Youngstrom, E. A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204-221.

Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(1), 85-93.

Table 1

Demographic Information for NSDUH 2010 (N = 24,120) and NSDUH 2013 (N = 23,627)

	NSDUH 2010		NSDUH 2010 - weighted		NSDUH 2013		NSDUH 2013 – weighted	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Sex								
Male	11550	47.89	77072108	50.70	11153	47.20	80958442	50.89
Race or ethnicity								
White	16548	68.61	110370000	72.60	15556	65.84	112110000	70.47
Black	2566	10.64	15353362	10.10	2760	11.68	16646754	10.46
Native American/Alaskan Native	352	1.46	545557.3	0.36	290	1.23	663428	0.42
Pacific Islander	98	0.41	382181.5	0.25	111	0.47	412202	0.26
Asian	804	3.33	5368092	3.53	842	3.56	6075838	3.82
Non-Hispanic multi-racial	632	2.62	1585125	1.04	699	2.96	2264940	1.42
Hispanic	3120	12.94	18427552	12.12	3369	14.26	20908594	13.14
Age, years								
21	2067	8.57	3747990	2.47	1896	8.02	3552987	2.23
22-23	3890	16.13	6933751	4.56	3907	16.54	7495927	4.71
24-25	3826	15.86	6819362	4.49	3843	16.27	7150236	4.49
26-29	2223	9.22	13235650	8.71	2098	8.88	14140246	8.89
30-34	2432	10.08	15665831	10.30	2248	9.51	15693018	9.86
35-49	6274	26.01	47134705	31.00	5695	24.10	45671241	28.71
50-64	2347	9.73	38228656	25.15	2663	11.27	41481434	26.08
65+	1061	4.40	20262186	13.33	1277	5.40	23898536	15.02
Marital Status								
Married	9817	40.70	83906698	55.19	9261	39.20	87212511	54.82
Widowed	489	2.03	7020001	4.62	483	2.04	6961520	4.38
Divorced/separated	2663	11.04	23233654	15.28	2584	10.94	23515748	14.78
Never married	11151	46.23	37867778	24.91	11299	47.82	41393846	26.02
Education Level								
Less than high school	2729	11.31	15964203	10.50	2458	10.40	14336157	9.01
High school	6768	28.06	41186190	27.09	6444	27.27	41818633	26.29
Some college	7351	30.48	41659029	27.40	7531	31.87	45314083	28.48
College or higher	7272	30.15	53218708	35.01	7194	30.45	57614753	36.22

Note. Demographics exclude those under age 18 and those who have not consumed at least one alcoholic beverage in the past year. Weighted refers to the application of NSDUH's person-level sampling weights to estimate the given statistics.

Table 2

Alcohol Use Disorder Criteria Endorsement Rates for NSDUH 2010 (N = 24,120) and NSDUH 2013 (N = 23,627)

	NSDUH 2010		NSDUH 2010 - weighted		NSDUH 2013		NSDUH 2013 – weighted	
Criterion	*n*	%	*N*	%	*n*	%	*n*	%
Withdrawal	613	2.54	3207272	2.11	599	2.54	2940331	1.85
Failure to Fulfill	704	2.92	2996331	1.97	589	2.49	2759042	1.73
Social/Interpersonal	885	3.67	3995752	2.63	756	3.20	3793623	2.38
Cut Down	947	3.93	5197930	3.42	885	3.75	5683733	3.57
Larger/Longer	1039	4.31	5202474	3.42	1046	4.43	5812842	3.65
Give Up	1062	4.40	4853596	3.19	965	4.08	5358323	3.37
Physical/Psychological	1260	5.22	6390363	4.20	1155	4.89	6101221	3.84
Hazardous Use	2190	9.08	9303113	6.12	1884	7.97	9395224	5.91
Tolerance	3891	16.13	18031896	11.86	3744	15.85	18860366	11.86
Time Spent	4209	17.45	19298414	12.69	3947	16.71	19857884	12.48

Note. Weighted refers to the application of NSDUH's person-level sampling weights to estimate the given statistics.

Table 3

Alcohol Use Disorder Diagnostic Rule Prevalence Rates For NSDUH 2010 (N = 24,120) and NSDUH 2013 (N = 23,627)

	NSDUH 2010		NSDUH 2010 - weighted		NSDUH 2013		NSDUH 2013 – weighted	
Diagnostic Rule	*n*	%	*n*	%	*n*	%	*N*	%
DSM-$5_{adjusted}$[a]	3816	15.82	17088854	11.24	3481	14.73	17556123	11.04
DSM-IV Abuse	1853	7.68	7815158	5.14	1558	6.59	8115700	5.10
DSM-IV Dependence	1495	6.20	7194747	4.73	1405	5.95	7056267	4.44
DSM-IV Alcohol Use Disorder	3348	13.88	15009905	9.87	2963	12.54	15171967	9.54
ICD-10 Dependence	1261	5.23	5991025	3.94	1191	5.04	5834651	3.67
Wakefield[a]	1546	6.41	7506921	4.94	1451	6.14	7892912	4.96
NSDUH 2010 Optimal Solution	2024	8.39	8883103	5.84	1856	7.86	9350704	5.88
NSDUH 2013 Optimal Solution	2205	9.14	9700828	6.38	2001	8.47	10165129	6.39
Overall Optimal Solution	2024	8.39	8883103	5.84	1856	7.86	9350704	5.88

Note. Weighted refers to the application of NSDUH's person-level sampling weights to estimate the given statistics. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; DSM-IV = Fourth Edition; ICD-10 = International Statistical Classification of Diseases and Related Health Problems, Tenth Edition; NSDUH = National Survey on Drug Use and Health. [a] = a variation of the diagnostic technique listed due to missing criteria within NSDUH.

Table 4

Overall Optimization Solution Results by Criterion Weighting Intervals

Criterion Weighting		Symptoms Considered in Optimization Procedure											
Consumption	Functional Impairment	Larger/ Longer	Cut Down	Time Spent	Failure to Fulfill	Social/ Interpersonal	Give Up	Hazardous Use	Physical/ Psychological	Tolerance	Withdrawal	Threshold	Out of
1	**0**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**3**	**9**
0.75	0.25	X	X	X	X	X		X	X	X	X	3	9
0.50	0.50	X	X	X	X	X		X	X	X	X	3	9
0.25	0.75	X	X	X		X	X	X	X	X		3	8
0	1	X	X	X		X	X	X	X	X		3	8

Note. The optimal solution referred to throughout the paper is presented in bold.

Table 5

Performance of Top Five Optimal Solutions with 100% Consumption as the Primary Optimization Criterion

Symptoms Considered in Optimization Procedure											Rule Characteristics		Rule Performance	
Rank Greater than or Equal to Base Rate	Larger / Longer	Cut Down	Time Spent	Failure to Fulfill	Social/ Interpersonal	Give Up	Hazardous Use	Physical/ Psychological	Tolerance	Withdrawal	Threshold	Set Size	Median Cohen's *d*	Range Cohen's *d*
1	X	X	X	X	X		X	X	X	X	3	9	0.58	(0.51, 0.69)
2		X	X		X		X			X	2	5	0.53	(0.50, 0.68)
3	X	X	X	X		X	X	X	X	X	3	9	0.59	(0.50, 0.68)
4	X	X	X		X	X	X	X	X	X	3	9	0.58	(0.52, 0.68)
5		X	X	X	X		X			X	2	6	0.53	(0.49, 0.67)

Note. Greater values of Cohen's *d* indicates a greater degree of separation between those diagnosing with alcohol use disorder versus those not diagnosing under the given criteria set and threshold.

Table 6

Adjusted Rand Index (ARI) and Kappa Estimates Across Diagnostic Rules in NSDUH 2013 (N = 23,627)

	Overall Optimal Solution	NESARC-III Optimal Solution	DSM-IV Dependence	DSM-IV AUD	DSM-$5_{adjusted}$	ICD-10 Dependence	Wakefield
Overall Optimal Solution	1	.75	0.80	0.70	0.60	0.73	0.60
NESARC-III Optimal Solution[a]	0.78	1	0.61	0.66	0.60	0.55	0.51
DSM-IV Dependence	0.82	0.86	1	0.56	0.48	0.90	0.62
DSM-IV AUD	0.75	0.57	0.61	1	0.66	0.49	0.48
DSM-$5_{adjusted}$[a]	0.66	0.50	0.54	0.73	1	0.41	0.45
ICD-10 Dependence	0.76	0.82	0.91	0.54	0.47	1	0.58
Wakefield Diagnosis[a]	0.63	0.66	0.65	0.53	0.51	0.61	1

Note. ARI estimates are on the upper diagonal and kappa estimates are on the lower diagonal. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; DSM-IV = Fourth Edition; AUD = Alcohol Use Disorder; ICD-10 = International Statistical Classification of Diseases, Tenth Edition. [a] = a variation of the diagnostic rule listed. Craving was excluded; however, agreement remains high within NESARC-III of the diagnostic rule listed with and without craving (Cohen's Kappa = 0.96).

Table 7

Phi Coefficient and Tetrachoric Correlation Estimates Across Diagnostic Rules in NSDUH 2013 (N = 23,627)

	Overall Optimal Solution	NESARC-III Optimal Solution	DSM-IV Dependence	DSM-IV AUD	DSM-$5_{adjusted}$	ICD-10 Dependence	Wakefield
Overall Optimal Solution	1	.80	0.83	0.77	0.70	0.78	0.64
NESARC-III Optimal Solution[a]	0.98	1	0.65	0.73	0.70	0.60	0.55
DSM-IV Dependence	0.99	0.99	1	0.66	0.60	0.92	0.65
DSM-IV AUD	1.00	1.00	1.00	1	0.73	0.61	0.57
DSM-$5_{adjusted}$[a]	1.00	1.00	1.00	0.94	1	0.55	0.57
ICD-10 Dependence	0.99	0.98	1.00	1.00	1.00	1	0.61
Wakefield[a]	0.91	0.92	0.91	0.89	0.91	0.89	1

Note. Phi coefficient estimates are on the upper diagonal and tetrachoric correlation estimates are on the lower diagonal. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; DSM-IV = Fourth Edition; AUD = Alcohol Use Disorder; ICD-10 = International Statistical Classification of Diseases, Tenth Edition. [a] = a variation of the diagnostic rule listed. Craving was excluded; however, agreement remains high within NESARC-III of the diagnostic rule listed with and without craving (Cohen's Kappa = 0.96).

Table 8

Weighted Odds Ratios (OR), Degree of Incremental Validity, and Chi-Square Difference Tests Adjusted for Sex and Age Comparing Various Diagnostic Rules with Overall and Second Overall Optimal Solutions in NSDUH 2013 (N = 23,627)

Diagnostic Rule	Formal treatment (e.g., outpatient)	Informal treatment (e.g., AA)	Mood disorder	Anxiety disorder	Suicide (i.e., thoughts, plan, or attempt)	Cannabis use disorder	Drug use disorder	Age of first drink ≤ 15 years old
	Odds Ratio (95% Confidence Interval)							
DSM-IV	9.80 (6.78, 14.15)	13.47 (8.19, 22.13)	3.52 (2.87, 4.31)	0.98 (2.18, 3.25)	3.10 (2.44, 3.94)	4.02 (3.11, 5.19)	7.72 (5.58, 10.68)	2.09 (1.82, 2.40)
DSM-$5_{adjusted}$[a]	8.67 (6.01, 15.51)	12.21 (7.41, 20.12)	3.22 (2.66, 3.90)	2.39 (1.96, 2.92)	3.02 (2.40, 3.81)	3.65 (2.85, 4.69)	5.64 (4.16, 7.64)	1.83 (1.61, 2.09)
ICD-10 Dependence	14.03 (9.47, 20.78)	21.05 (12.98, 34.15)	4.28 (3.32, 5.52)	3.48 (2.71, 4.48)	4.13 (3.13, 5.46)	3.73 (2.62, 5.31)	7.69 (5.44, 10.88)	2.40 (1.97, 2.93)
Wakefield[a]	11.71 (8.08, 16.98)	19.30 (12.08, 30.85)	4.54 (3.53, 5.62)	2.91 (2.31, 3.66)	3.93 (3.02, 5.14)	4.01 (2.92, 5.50)	6.18 (4.44, 8.61)	2.18 (1.82, 2.62)
Overall Optimal Solution	12.20 (8.38, 17.75)	17.52 (10.79, 28.45)	4.39 (3.48, 5.52)	2.92 (2.35, 3.63)	4.12 (3.16, 5.37)	4.40 (3.31, 5.84)	8.10 (5.96, 11.01)	2.44 (2.06, 2.88)
Second Overall Optimal Solution	11.83 (8.10, 17.28)	18.48 (11.29, 30.22)	3.77 (3.01, 4.71)	2.62 (2.10, 3.27)	3.20 (2.50, 4.08)	4.22 (3.19, 5.58)	6.57 (4.79, 9.02)	2.23 (1.88, 2.64)

	Degree of Incremental Validity (Δc) and Chi-Square Difference Test															
	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}	Δc	χ^2_{diff}
Incremental validity over DSM-IV																
Overall Optimal Solution	0.009	260655.60*	0.011	233424.36*	0.003	293354.67*	0.005	59187.53*	0.003	298199.13*	0.000	58617.65*	0.002	117755.49*	0.000	187823.26*
Second Overall Optimal Solution	0.008	259621.39*	0.013	293722.40*	0.001	101322.54*	0.003	14285.43*	0.002	39525.62*	0.001	53306.12*	0.002	33302.77*	0.001	78390.85*
Incremental validity over DSM-$5_{adjusted}$																
Overall Optimal Solution	0.012	392164.93*	0.017	318936.22*	0.005	439122.72*	0.007	149098.27*	0.003	309550.72*	0.001	111370.82*	0.005	385282.84*	0.002	452821.05*
Second Overall Optimal Solution	0.023	348167.01*	0.019	353622.76*	0.001	166086.39*	0.003	52057.19*	0.002	44699.71*	0.002	85972.58*	0.006	161391.75*	0.001	227207.52*
Incremental validity over ICD-10																
Overall Optimal Solution	0.021	297258.23*	0.008	165365.16*	0.011	766089.27*	0.004	129690.65*	0.011	436678.75*	0.015	363412.52*	0.026	537140.76*	0.009	625973.93*
Second Overall Optimal Solution	0.031	413246.37*	0.034	337417.06*	0.009	478050.73*	0.006	108709.11*	0.009	143532.35*	0.017	332001.54*	0.026	332694.80*	0.009	453818.50*
Incremental validity over Wakefield																
Overall Optimal Solution	0.033	430893.07*	-0.003	219689.35*	0.005	509419.76*	0.007	242056.7*	0.008	383263.98*	0.007	257905.18*	0.025	691473.25*	0.009	641859.23*
Second Overall Optimal Solution	0.030	491826.29*	0.016	348568.10*	0.004	321577.46*	0.006	158202.31*	0.006	125849.94*	0.010	255682.58*	0.022	433543.62*	0.008	467412.21*

Note. Weighted refers to the application of NSDUH's person-level sampling weights to estimate the given statistics. [a] = a variation of the diagnostic technique listed due to missing criteria within NSDUH. The Chi-Square difference test is a one degree of freedom test. [a] = a variation of the diagnostic rule listed. *$p < .001$