Published in final edited form as:

Addict Behav. 2019 July ; 94: 57–64. doi:10.1016/j.addbeh.2019.02.005.

Toward more efficient diagnostic criteria sets and rules: The use of optimization approaches in addiction science

Jordan E. Stevens^{a,*}, Douglas Steinley^a, Yoanna E. McDowell^a, Cassandra L. Boness^a, Timothy J. Trull^a, Christopher S. Martin^b, and Kenneth J. Sher^a

^aDepartment of Psychological Science, University of Missouri, United States

^bDepartment of Psychological Science, University of Pittsburgh, United States

Abstract

Psychiatric diagnostic systems, such as *The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5), use expert consensus to determine diagnostic criteria sets and rules (DCSRs), rather than exploiting empirical techniques to arrive at optimal solutions (OS). Our project utilizes complete enumeration (i.e., generating all possible subsets of item combinations *A* and *B* with all possible thresholds, *T*) to evaluate all possible DCSR given a set of relevant diagnostic data. This method yields the entire population distribution of diagnostic classifications (i.e., diagnosis of the disorder versus no diagnosis) produced by a set of dichotomous predictors (i.e., diagnostic criteria). Once unique sets are enumerated, optimization on some predefined correlate or predictor will maximally separate diagnostic groups on one or more, disorder-specific “outcome” criteria. We used this approach to illustrate how to create a common Substance Use Disorder (SUD) DCSR that is applicable to multiple substances. We demonstrate the utility of this approach with respect to alcohol use disorder and Cannabis Use Disorder (CUD) using DSM-5 criteria as input variables. The optimal SUD solution with a moderate or above severity grading included four criteria (i.e. 1) having a strong urge or craving for the substance (CR), 2) failure to fulfill major role obligations at work school or home (FF), 3) continued use of the substance despite social or interpersonal problems caused by the substance use (SI) and 4) physically hazardous use (HU)) with a diagnostic threshold of two. The derived DCSR was validated with known correlates of SUD and performed as well as DSM-5. Our findings illustrate the value of using an empirical approach to what is typically a subjective process of choosing criteria and algorithms that is prone to bias. The optimization of diagnostic criteria can reduce criteria set sizes, resulting in decreased research, clinician, and patient burden.

*Corresponding author at: Department of Psychological Science, Psychology Building, 200 South Seventh Street, Columbia, MO 65211, United States. jes9bc@mail.missouri.edu (J.E. Stevens).

Contributors

Authors Stevens and Steinley developed and performed the statistical techniques described in this manuscript. Stevens wrote the first draft of the manuscript and all other authors (McDowell, Boness, Trull, Martin, Steinley, and Sher) contributed to and have approved the final manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Keywords

Diagnosis; Optimization; Alcohol Use Disorder; Cannabis Use Disorder; Substance Use Disorder; Classification

1. Introduction

Valid and reliable diagnostic criteria sets and rules (DCSRs) are fundamental to clinical research and practice. Current diagnostic standards, such as the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5; American Psychiatric Association [APA], 2013) utilize expert consensus informed by relevant literature to develop DCSR of psychopathology (Frances & Widiger, 2012; Hasin et al., 2013). For example, in revising the DSM, the DSM-5 Substance-Related Disorders Workgroup (Hasin et al., 2013) tasked itself with answering questions such as: “Which if any diagnostic criteria can be dropped?” and “What should the diagnostic threshold be?” (Hasin et al., 2013). The ability to answer questions such as these is limited by the inherent subjectivity that permeates the consensus process (Frances & Widiger, 2012; Wakefield, 2015). However, statistical techniques can be used to empirically derive new DCSR where the process is transparent and biases are made explicit.

The optimization procedure described here (Steinley et al., 2016, b; Stevens et al., 2018), can empirically derive DCSR based on a priori clinical correlates and outcomes that serve as optimization criteria (OptCrit). Raffo and colleagues (2018) have successfully demonstrated that a similar method could be used to create short-forms of diagnosis by optimizing the correlations between the full DSM-5 Alcohol Use Disorder (AUD) criteria set with diagnostic thresholds (i.e., the number of criteria in a set one must endorse to diagnose) of one or two. By optimizing on the basis of part-whole correlations, Raffo et al.’s approach is likely influenced by intrinsic correlated error, a limitation not shared with our current approach. Additionally, the Raffo et al., (2018) approach employed a post hoc evaluation of a fixed number of diagnostic thresholds subsequent to short-form derivation, in effect, evaluating only a restricted subset of possible criteria sets and diagnostic rules and thus not guaranteed to identify an optimal solution (OS). The current procedure, however, is optimizing on external criteria to derive novel diagnoses rather than creating a shortened form of the diagnostic sets. This approach uses complete enumeration to compare all possible criteria set sizes and diagnostic thresholds from a selected set of criteria. Table 1 illustrates complete enumeration on an 11 item criteria set varying the diagnostic threshold from 1 to 11, producing 11,264 DCSR. Once completely enumerated, the distribution of all measures of separation produced can be used to identify where the diagnostic grouping (i.e., those diagnosing under the given rule versus those not diagnosed by the rule) are meaningfully distinguishable using some objective function, such as maximizing the effect size. Here, we extend our previous work (Steinley et al., 2016, b; Stevens et al., 2018) by deriving a common DCSR across alcohol and cannabis in order to illustrate how this approach can be used to identify a single OS common to alcohol and other drugs of abuse (e.g., Budney, 2006). Although the focus of the current paper is to describe how optimal DCSR can be derived, applications using the current approach can be found in optimizing

the diagnosis of Alcohol Use Disorder (AUD; Boness et al., 2018), resulting in variable, equally performing diagnostic rules (i.e., set size of 9 with a diagnostic threshold of three or a set size of five with a diagnostic threshold of 2).

2. Method

Below we describe the process that one would go through in using complete enumeration to identify an optimal DCSR, with the application provided in Section 4. This includes both basic decisions about selection of datasets, OptCrit, and external validation approaches.

2.1. Steps for optimizing diagnosis

Step 0: Data preparation and a priori decisions.—In order to use the proposed method, there are a number of factors to consider a priori including:

- *Diagnostic Criteria (DxCrit).* The DxCrit considered for optimization should have a strong theoretical and/or empirical basis.¹ As is common practice, DxCrit should be dichotomously coded, where a ‘1’ indicates the presence of a symptom and a ‘0’ indicates its absence.
- *Optimization criteria (OptCrit).* One or more indicators can be used as the OptCrit. The OptCrit serve as the basis for selecting where DCSRs maximally differentiate diagnostic groupings. The OptCrit must be correlated and be a valid marker of the disorder of interest. In the case of optimizing across multiple OptCrit simultaneously, each OptCrit should be given a weighting after being scaled to a common metric.
- *Constraints.* There are numerous constraints that one should consider before proceeding with the optimization. One general class of constraints are those involving the types of candidate OSs that should be considered. For example, it is known that a diagnostic threshold of one is problematic due to false positive rates (Hasin et al., 2013; Martin, Chung, & Langenbucher, 2008), so it may be beneficial to only examine sets with a threshold of two or greater. In some instances, it also may be desirable to limit candidate OSs to only those including/excluding one or more specific criteria based on theoretical properties. Similarly, an a priori target prevalence can be established if deemed desirable.

Step 1: Randomly divide dataset into k non-overlapping folds.—Datasets are randomly divided into k , roughly equivalent, non-overlapping subsets. Splitting the data this way is a cross-validation method used in classification to safeguard against overfitting to the full dataset (Rodriguez, Perez, & Lozano, 2010; Stone, 1974). The optimization procedure occurs within each fold independently, allowing for examination of each DCSR within subsets of the data. If sample sizes allow, 5 to 10 folds are typically sufficient (Rodriguez et al., 2010).

¹Although complete enumeration is designed to reduce the set size when possible, this is only computationally feasible with approximately 25 or less DxCrit on a desktop computer, so thoughtful consideration should be used for inclusiveness.

Step 2: Complete enumeration of diagnostic rules.—For all C symptoms that are selected to be included in the optimization procedure, A symptoms that predict the OptCrit and B symptoms that do not provide additional prediction of the OptCrit are identified. For a set of C symptoms, the number of sets considered will consist of $2^C - 1$ possible subsets (excluding the empty set) when the diagnostic threshold is one. When all possible DCSRs are enumerated, the number of rules to examine becomes immense. For the simplest instance of three DxCrit assessed (x, y, z), full enumeration would result in 12 possible rules: (i) endorsing x , (ii) endorsing y , (iii) endorsing z , (iv) endorsing either x or y , (v) endorsing either x or z , (vi) endorsing either y or z , (vii) endorsing either x, y , or z , (viii) endorsing both x and y , (ix) endorsing both x and z , (x) endorsing both y and z , (xi) endorsing two out of x, y , and z , and (xii) endorsing x, y , and z .

Step 3: Calculate distance between diagnostic groups.—A predefined objective function is used to determine the magnitude of the group differences between those diagnosing under each diagnostic rule versus those not diagnosing. The objective function could be any measure of group separation, such as Cohen's d or Mahalanobis' distance.

Step 4: Rank performances of diagnostic sets.—Within each fold, each diagnostic rule is assigned a rank in accordance to the performance on separating the diagnostic groupings. A rank of 1 will be assigned to the diagnostic rule with the maximal separation between diagnostic groupings, whereas, high ranks will be assigned to rules that fail to differentiate between those diagnosing under the rule versus those not diagnosing. Rankings are assigned in this step instead of using the raw estimates of the distance measure to ensure that the results from a single fold do not determine the OS.

Step 5: Selection of optimal set.—The ranking of each diagnostic rule is averaged across the k folds to give an indication of its overall performance. The overall optimal rule will meet all pre-defined constraints and have the lowest average ranking across folds. In order to choose DCSRs that are valid for both alcohol and cannabis, the rank ordering of different DCSRs is then averaged across alcohol and cannabis. The DCSR with the lowest rank ordering across drugs is selected as the optimal cross-drug diagnostic rule.

Step 6: External validation.—The optimal DCSR should be validated on a set of known correlates of the disorder that did not serve as DxCrit or OptCrit to examine if the rule performs at least as well as established alternative diagnostic schemes.

3. Measures

3.1. Dataset

Data were drawn from the National Epidemiological Survey on Alcoholism and Related Conditions-III (NESARC-III; Grant et al., 2009). NESARC-III, conducted from 2012 to 2013, was sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Multistage probability sampling (see Grant et al., 2015) was used to randomly select a nationally representative sample of United States citizens at least 18 years old. There were a total of 36,309 participants in the NESARC-III survey, however, the current sample was restricted to individuals who were considered to be current drinkers (determined by having

at least 12 drinks in the past year) or current cannabis users (indicated in NESARC-III as those using cannabis at least once in the past year). The final sample sizes were 20,495 participants who have had at least 12 drinks in the past year and 3701 participants who have used cannabis in the past year. (Note, the optimization procedure does not utilize sampling weights, although our external validation procedures do, see below).

3.1.1. DCSRs for AUD and CUD—DxCrit for AUD and Cannabis Use Disorder (CUD) within the DSM-5 each consist of 11 criteria: 1) using a substance in larger amounts or over a longer period of time (LL), 2) attempting/wanting to cut down (CD), 3) spending a great deal of time obtaining/using/recovering from effects (TS), 4) having a strong urge/craving (CR), 5) failure to fulfill major role obligations at work school or home (FF), 6) continued use despite social or interpersonal problems (SI), 7) giving up important social/occupational/recreational activities (GU), 8) physically hazardous use (HU), 9) continued use despite persistent physical or psychological problems (PP), 10) tolerance (TL), and 11) withdrawal (WD).

Parallel items across AUD and CUD DxCrit (excluding withdrawal) were used to construct our overall SUD solution. Table 2 includes the prevalence rates for the DxCrit within the current sample. While many of the characteristic withdrawal symptoms are overlapping for AUD and CUD (e.g., requiring cessation after prolonged use of the substance), there are specific criteria for each due to the differing substance-specific physiological changes that occur (APA, 2013). Consequently, sub-stance-specific criteria were used for each disorder.

3.1.2. Optimization criteria (OptCrit)—When selecting the OptCrit, one should either select a single criterion strongly associated with the diagnosis, or multiple criteria moderately related to the disorder. Specific to SUD, research has shown that heavy use over time is the most parsimonious construct for explaining the various physical and social consequences that occur within the dependence syndrome (Martin, Langenbucher, Chung, & Sher, 2014; Rehm et al., 2013; Rehm & Roerecke, 2013). For example, alcohol consumption has been found to have a strong monotonic association with the DSM-5 criterion count (Dawson, Saha, & Grant, 2010; Lane & Sher, 2014; Saha, Stinson, & Grant, 2007). Consequently, composites of alcohol and cannabis consumption were selected as the OptCrit for deriving a general SUD diagnosis.

Alcohol Consumption. Alcohol consumption was assessed using a composite measure of past-year drinking measures including: maximum number of drinks in a single day, frequency of drinking, typical quantity of specific alcoholic beverages, frequency of consuming the maximum number of reported drinks, binge frequency (i.e., 4 or more drinks for women; 5 or more drinks for men), frequency of intoxication, and exceeding daily limits defined by 4 or more drinks for women in a single day and 5 or more drinks for men in a single day (NIAAA, 2008; $\alpha = 0.75$). To account for sex differences in consumption (Baraona et al., 2001; Thomasson, 2002; Wilsnack, Vogeltanz, Wilsnack, & Harris, 2000; York & Welte, 1994), alcohol consumption variables were standardized by sex *Cannabis consumption*. As other studies have found, quantifying cannabis consumption can be difficult due to the lack of objective measures of the amount or the potency of cannabis (Compton, Saha, Conway, & Grant, 2009; Zeisser et al., 2012). NESARC-III assessed

quantity of consumption by the number of joints smoked, not including any measure of potency. The inclusion of the NESARC-III quantity measure was found to decrease the overall alpha of the composite ($\alpha = 0.55$ for frequency alone to $\alpha = 0.44$ with quantity). For this reason, a standardized composite consisting of past-year frequency of use and past-year frequency of maximum use (rescored to weekly use) was utilized as the OptCrit.

3.1.3. External validators—The overall optimal SUD rule was validated using known correlates of at least one of the disorders. External validators were dichotomously coded with ‘1’ indicating the presence of the correlate within the past year (except where indicated), whereas a ‘0’ indicated the absence. External validators considered were: alcohol and cannabis treatment seeking behavior (e.g., detoxification ward or clinic, rehabilitation program, alcoholics anonymous or narcotics anonymous for cannabis use); past year DSM-5 tobacco use disorder, comorbid substance use disorder; first drink before age 16; first cannabis use before age 16; parental problem-drinking; parental (any) problem-drug use; mood disorder (e.g., major depressive episode); past year anxiety disorder (e.g., panic disorder); past year DSM-5 antisocial personality disorder with impairment, any lifetime personality disorder (e.g., borderline personality disorder); and lifetime suicide attempt within those individuals endorsing symptoms under the classification of a mood disorder.

4. Demonstration

In this section, we provide a detailed, step-by-step example of optimizing diagnosis across AUD and CUD to develop a general SUD diagnosis at a moderate or above severity level (i.e., endorsing at least four of the 11 criteria) defined by the DSM-5.

Step 0: Data preparation and a priori decisions

- *Diagnostic Criteria (DxCrit)*. The DxCrit used to develop an overall SUD diagnosis are the 11 DSM-5 AUD and CUD criteria. In total, 22 DxCrit (11 from each substance) were used.
- *Optimization criterion (OptCrit)*. The OptCrit selected were alcohol and cannabis consumption composites. Each criterion was optimized on independently for the respective disorder.
- *Constraints*. Due to issues with false positive rates (Hasin et al., 2013), Ods with a threshold less than two were not considered as eligible candidate optimal solution. Optimization was completed on current alcohol users and current cannabis users. Minimum prevalence rates of both AUD and CUD were set to the weighted average past year prevalence rate of a moderate or severe DSM-5 diagnosis within NESARC-III. The moderate or severe base rate was used since studies indicate that moderate DSM-5 is a much more reasonable estimate of what we expect from ‘true’ SUD (Martin et al., 2011, b). For this example, the moderate to severe base rate was set at 11.74% (average of the 11.29% AUD moderate to severe prevalence rate and the 12.18% CUD prevalence rate).

Step 1: Randomly divide dataset into k non-overlapping folds.

Each of the datasets were randomly divided into 5 non-overlapping folds, for a total of ten folds (5 folds within the AUD dataset and 5 folds within the CUD dataset).

Step 2: Complete enumeration of diagnostic rules.

For a diagnosis of either AUD or CUD, there are 11 DxCrit identified by the DSM-5 that were considered. The number of sets considered with a threshold of one were $2^{11}-1 = 2047$ (although these were examined for completion, they were not eligible for consideration as the OS, as described previously). Table 1 presents the full enumeration of all DxCrit when the threshold is varied from 1 to 11. In total, there were 11,264 possible diagnostic rules enumerated for examination.

Step 3: Calculate distance between diagnostic groups.

Cohen's d was used to determine the differences in consumption rates of those diagnosing under each diagnostic rule and those not diagnosing within each disorder. For example, if a set size was 6 with a threshold of 3, participants endorsing < 3 DxCrit would be in the non-diagnostic group, while those endorsing 3 or more symptoms would be in the diagnostic group. An effect size of the difference in optimization criterion between diagnostic groups was then calculated within each fold of the dataset.

Step 4: Rank performances of diagnostic rules.

Within each fold, ranks were assigned to each Cohen's d value. Greater degrees of separation are indicated by high Cohen's d values, and were assigned low ranks and vice versa.

Step 5: Selection of optimal set.

The ranking of each diagnostic rule was averaged across all ten folds. The diagnostic rule with the minimum ranking, that also diagnosed at least the minimum base rate within at least one fold of the dataset was identified as the candidate OS.

Step 6: External validation.

AUD and CUD diagnoses were constructed separately with their respective DxCrit, and jointly, such that diagnosis was based on a diagnosis under the AUD DxCrit AND/OR CUD DxCrit (i.e., meeting criteria for AUD and/or CUD constituted a positive SUD diagnosis). In total, the OS was assessed in three ways: 1) optimal DCSR based on AUD DxCrit, 2) optimal DCSR based on CUD DxCrit, and 3) optimal DCSR based on meeting DxCrit within AUD OR CUD. Similarly, the OS was compared to moderate-severe DSM-5 AUD and CUD diagnoses (i.e., a past year criterion of four or more because, as we have argued else- where, we strongly believe the DSM-5 threshold is too low and produces too many false positive diagnoses; Martin et al., 2011, b). This was done separately for each substance as well as a combined DSM-5 SUD diagnosis consisting of an AUD OR CUD diagnoses.

5. Results and discussion

5.1. Selection of optimal sets

The identified moderate-severe optimal diagnostic rule included a set size of 4 with a diagnostic threshold of 2. The DxCrit of the derived general SUD rule are: 1) having a strong urge or craving for the substance (CR), 2) failure to fulfill major role obligations at work school or home (FF), 3) continued use of the substance despite social or inter-personal problems caused by the substance use (SI) and 4) physically hazardous use (HU).

5.2. Validation

Planned comparisons incorporating NESARC-III's complex sampling design were used to externally validate the optimal SUD solution. To assess the performance of the SUD optimal diagnostic set compared to DSM-5 AUD and/or CUD diagnoses, Proc Surveylogistic in SAS was used to obtain odds ratios (OR) of each of the diagnostic techniques alone (top half of Table 3), along with the incremental validity of adding in the OS to predictions where the DSM-5 diagnosis is already in the model (bottom half of Table 3). In each of these analyses, age and sex were included as covariates. The degree of incremental validity over the DSM-5 diagnoses is included examining the change in concordance (c) and results from the χ^2 difference test between the models with and without the moderate-severe OS.

The logistic regression predicting the external validators indicate the SUD OS performs equally as well as the DSM-5 SUD diagnosis. The c 's reported in Table 3 are small, indicating that our SUD OS has modest incremental validity over DSM-5; however, any incremental validity is viewed as a success since seven fewer symptoms are needed to assess diagnosis. This suggests our SUD OS may be much more efficient without significant loss of information. Tables 4 includes all ARI (upper diagonal; Steinley, Brusco, & Hubert, 2016) and simple Kappa (lower diagonal; Viera & Garrett, 2005) values for all diagnostic schemes examined in this paper. These estimates indicate good agreement between the derived general SUD and the DSM-5 SUD diagnosis (ARI = 0.67; Kappa = 0.72). Table 5 includes correlational information between all diagnostic schemes examined. High correlations were found between the derived SUD diagnosis and the DSM-5 SUD diagnosis ($\phi = 0.72$; tetrachoric = 0.94).

5.3. Limitations and extensions

The proposed optimization method effectively derives optimal diagnostic rules, providing an empirical tool to refine DxCrit sets. However, this refinement comes at a cost. Complete enumeration of a diagnostic set takes great computational power, especially with large item sets. Although the approach can easily handle the 11 item DxCrit set used in the demonstration, computational limits arising from von Neumann's law (1966) will greatly slow down the procedure with in-creasing set sizes. Set sizes with greater numbers of symptoms can utilize metaheuristic optimization techniques (Holland, 1975; Kirkpatrick et al., 1986; Glover, 1986) to identify diagnostic rules, however, not all of these procedures guarantee a globally optimal rule as is guaranteed by the current method.

The current optimization approach is illustrated in the context of deriving optimal SUD diagnoses that are generalizable across substances. However, this approach could be used for any set of dichotomous predictors (with or without a threshold) to summarize some outcome variable. While deriving diagnostic rules is one such application, there are a number of assessment settings where this method could be extended. For example, in personality assessment we could use (with minor adjustments) the same procedure to: (1) determine the number of items in a personality test that can be missing while still providing good estimates of full scale means, (2) develop optimal short forms, and (3) identify test items that are potentially inflating personality-outcome correlations because of spurious predictor-criterion contamination. Basically, complete enumeration provides a disciplined approach for exhaustively evaluating all possible states of nature in order to identify variables and rules for combining them that best meet a specific goal. Application of such an approach is potentially wide ranging.

However, in any application, the optimization procedure relies on candidate DxCrit and OptCrit that have a theoretical basis since the procedure is blind to the quality of the variables, as with any assessment object instrument (Watson, 2012). Any minor adjustments to these variables could lead to substantially different solutions since the goal of the procedure is just to find the most efficient diagnostic set given constraints (e.g., diagnostic base rate), a criteria set to optimize (e.g., symptoms reflecting given disorder) and one or more optimization criteria (e.g., variables that are associated with the disorder that are not used as symptoms of the disorder).

Although factor analyses tend to reveal a unidimensional structure underlying AUD criteria (Hasin et al., 2013), scrutiny of the consistency of item response theory (IRT) thresholds across more than two dozen studies illustrate critical, instrument-specific operationalizations of criteria are (see Lane & Sher, 2014) leading us to be cautious in overly interpreting findings based on any single data base, even one as representative and large as NESARC-III. The goal of the paper was simply to highlight that empirical tools can be used to refine diagnosis. The approach demonstrates that using data driven methods, a diagnostic rule with consistent criteria across multiple substances can be identified. There has been no attempt to identify subfactors because there is not a clear mapping of specific criteria onto underlying basic addiction constructs (Bickel, Crabbe, & Sher, 2019; Martin et al., 2014).

Although the DSM-5 SUD criteria were used to derive a diagnostic rule, we would never argue that these criteria are the best symptoms for the derivation described here. Rather, this approach is simply a tool that experts can use to identify the most efficient set given the selected DxCrit. Unfortunately, most large, representative data sets have a restricted number of relevant candidate criteria to evaluate. Future data collection efforts should be more inclusive with respect to candidate criteria and, in our opinion, more translational in focus and related to major theories of addiction.

With these caveats in mind, it is instructive to point out that the optimal criteria set included items that were both relatively severe (e.g., failure to fulfill major role obligations), moderate (e.g., continued use despite interpersonal problems), and relatively mild (e.g., craving) from an IRT perspective suggesting broad “trait coverage.” Craving is a central concept in many

theories of addiction (e.g., “incentive sensitization,” Berridge & Robinson, 2016). Continued use despite interpersonal problems can be viewed as reflecting, in part, compulsion to use (i.e., use in the face of punishment)...a key manifestation of addiction (Piazza & Deroche-Gamonet, 2014). Failure to fulfill major role obligations can be construed as reflecting a drug “hijacking” reward systems so that substance use becomes more important than other natural reinforcers in the user’s environment (Volkow & Li, 2005). In our opinion, hazardous use can be reflective of compulsion (e.g., choosing to use with the knowledge that such use could cause imminent harm) but could also indicate a general pattern of heedless behavior (Martin et al., 2011, b); that is, the criterion is intrinsically heterogeneous and it is difficult to determine the extent it is reflecting an addictive (i.e., compulsive) or impulsive phenomenon.

6. Conclusions

This paper provides a guide for a data-driven, empirical optimization procedure that can be used for diagnostic refinement. The results from the example, creating a diagnostic set to optimally diagnose AUD or CUD, demonstrated that large sets can be reduced to more efficient and manageable set sizes to decrease research, clinician, and patient burden during assessment. In the current case we identified 4 items: 1) having a strong urge or craving for the substance (CR), 2) failure to fulfill major role obligations at work school or home (FF), 3) continued use of the substance despite social or interpersonal problems caused by the substance use (SI) and 4) physically hazardous use (HU). The first three of these criteria represent aspects of the dependence syndrome (i.e., craving) as well as functional impairment arising from substance use (i.e., failure to fulfill and continued to use despite social or inter-personal problems). In contrast, hazardous use is a somewhat complex criterion in that it can reflect simply heedless behavior or an aspect of compulsive use. This highlights that the described procedure can draw variance that accounts for a larger criteria set within a single disorder and across multiple disorders. Future nosologists will have to weigh the cost and benefits of unified criteria sets versus those that are customized for specific SUDs.

Derived diagnostic sets can be compared to current diagnostic standards, allowing for an objective, less biased measurement development tool. Moreover, we extended our methodology for the current paper to show how the technique can be applied to a challenging problem in addiction science; *how can we best identify criteria for SUDs that work well across multiple substances*. In the current paper we focused on the two psychoactive SUDs with the highest prevalence rates in the population; however, there is no reason this approach cannot be extended to include additional substances.

Acknowledgments

The present study was supported by the NIH grants T32AA013526, F31AA026177, R01 AA13397, K24 AA020840 and R01AA024133.

Role of funding sources

Funding for this study was provided by the NIH grants T32AA013526, F31AA026177, R01 AA13397, K24 AA020840 and R01AA024133. This is a secondary analysis of the NESARC-III dataset that was conducted by the

entramural branch of NIAAA. The authors were funded in part by NIH grants T32AA013526, F31AA026177, R01AA13397, K24 AA020840 and R01AA024133. Consequently, NIAAA is responsible for the conduct of NEARC-III but the investigators were wholly responsible for the analyses, interpretation of the data, and the writing presented in this manuscript.

References

- American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th Ed.). Arlington, VA: American Psychiatric Publishing.
- Baraona E, Abittan CS, Dohmen K, Moretti M, Pozzato G, Chayes ZW, ... Lieber CS (2001). Gender differences in pharmacokinetics of alcohol. *Alcoholism: Clinical and Experimental Research*, 25(4), 502–507.
- Berridge KC, & Robinson TE (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, 71(8), 670–679. [PubMed: 27977239]
- Bickel WK, Crabbe JC, & Sher KJ (2019). What Is Addiction? How Can Animal and Human Research Be Used to Advance Research, Diagnosis, and Treatment of Alcohol and Other Substance Use Disorders? *Alcoholism: Clinical and Experimental Research*, 43(1), 6–21.
- Boness CL, Stevens JE, Steinley D, Trull T, & Sher KJ (2018). Deriving alternative criteria sets for alcohol use disorders using statistical optimization: Results from the National Survey on Drug Use and Health. *Experimental and clinical psycho-pharmacology*. 10.1037/pha0000249.
- Budney AJ (2006). Are specific dependence criteria necessary for different substances: how can research on cannabis inform this issue? *Addiction*, 101, 125–133. [PubMed: 16930169]
- Compton WM, Saha TD, Conway KP, & Grant BF (2009). The role of cannabis use within a dimensional approach to cannabis use disorders. *Drug & Alcohol Dependence*, 100(3), 221–227. 10.1016/j.drugalcdep.2008.10.009. [PubMed: 19062204]
- Dawson DA, Saha TD, & Grant BF (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. [PubMed: 19782481]
- Frances AJ, & 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.
- Glover F (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13(5), 533–549.
- Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, & Pickering R (2003). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADISIV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug and Alcohol Dependence*, 71(1), 7–16. [PubMed: 12821201]
- Grant JD, Agrawal A, Bucholz KK, Madden PA, Pergadia ML, Nelson EC, ... Whitfield JB (2009). Alcohol consumption indices of genetic risk for alcohol dependence. *Biological Psychiatry*, 66(8), 795–800. [PubMed: 19576574]
- Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, ... Hasin DS (2015). Epidemiology of DSM-5 alcohol use disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*, 72(8), 757–766. [PubMed: 26039070]
- Hasin DS, & Grant BF (2015). The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. *Social Psychiatry and Psychiatric Epidemiology*, 50(11), 1609–1640. [PubMed: 26210739]
- Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, ... Schuckit M (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry*, 170(8), 834–851. [PubMed: 23903334]
- Holland JH (1975). *Adaptation in natural and artificial systems An introductory analysis with application to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.
- Kirkpatrick S, Gelatt CD, & Vecchi MP (1986). Optimization by simulated annealing. *World Scientific Lecture Notes in Physics Spin Glass Theory and Beyond*, 339–348. 10.1142/9789812799371_0035.

- Lane SP, & Sher KJ (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. [PubMed: 26783505]
- Martin CS, Chung T, & Langenbucher JW (2008). How should we revise diagnostic criteria for substance use disorders in the DSM-V? *Journal of Abnormal Psychology*, 117(3), 561–575. [PubMed: 18729609]
- Martin CS, Langenbucher JW, Chung T, & Sher KJ (2014). Truth or consequences in the diagnosis of substance use disorders. *Addiction*, 109(11), 1773–1778. [PubMed: 24913314]
- Martin CS, Sher KJ, & 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. [PubMed: 21683051]
- Martin CS, Steinley DL, Vergès A, & Sher KJ (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. [PubMed: 21557890]
- National Institute on Alcohol Abuse and Alcoholism (2008). Wave 2 NESARC data notes. Rockville, MD: Author.
- Piazza PV, & Deroche-Gamonet V (2014). A general theory of transition to addiction it was and a general theory of transition to addiction it is. *Psychopharmacology*, 231(19), 3929–3937. [PubMed: 24908270]
- Raffo CD, Hasin DS, Appelbaum P, & Wall MM (2018). A data-driven method for identifying shorter symptom criteria sets: the case for DSM-5 alcohol use disorder. *Psychological Medicine*, 1–9.
- Rehm J, Marmet S, Anderson P, Gual A, Kraus L, Nutt DJ, Room R, Samokhvalov AV, ... Wiers RW, Gmel G (2013). Defining substance use disorders: Do we really need more than heavy use? *Alcohol and Alcoholism*, 48, 633–640. [PubMed: 23926213]
- Rehm J, & Roerecke M (2013). Reduction of drinking in problem drinkers and all-cause mortality. *Alcohol and Alcoholism*, 48, 509–513. [PubMed: 23531718]
- Rodriguez JD, Perez A, & Lozano JA (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569–575. [PubMed: 20075479]
- Saha TD, Stinson FS, & Grant BF (2007). The role of alcohol consumption in future classifications of alcohol use disorders. *Drug and Alcohol Dependence*, 89, 82–92. [PubMed: 17240085]
- Steinley D, Brusco MJ, & Hubert L (2016). The variance of the adjusted Rand index. *Psychological Methods*, 21(2), 261–272. [PubMed: 26881693]
- Steinley D, Lane SP, & Sher KJ (2016). Determining optimal diagnostic criteria through chronicity and comorbidity. *In Silico Pharmacology*, 4(1), 1–12. [PubMed: 26831872]
- Stevens JE, Steinley D, Boness CL, Trull T, Wood P, & Sher KJ (2018). Combinatorial optimization of classification decisions: An application to refine psychiatric diagnoses. [10.31234/osf.io/jpnmf](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.
- Thomasson HR (2002). Gender differences in alcohol metabolism Recent developments in alcoholism (pp. 163–179). Boston, MA: Springer.
- Viera AJ, & Garrett JM (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360–363. [PubMed: 15883903]
- Volkow N, & Li TK (2005). The neuroscience of addiction. *Nature Neuroscience*, 8(11), 1429–1430. [PubMed: 16251981]
- Von Neumann J, & Burks AW (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1), 3–14.
- Wakefield JC (2015). DSM-5 substance use disorder: How conceptual missteps weakened the foundations of the addictive disorders field. *Acta Psychiatrica Scandinavica*, 132(5), 327–334. [PubMed: 25968656]
- Watson D (2012). Objective tests as instruments of psychological theory and research Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, & Sher KJ (Eds.). *APA Handbook of Research Methods in Psychology, Vol. 1 Foundations, Planning, Measures, and Psychometrics*, 349–369. [10.1037/13619-019](https://doi.org/10.1037/13619-019).

- Wilsnack RW, Vogeltanz ND, Wilsnack SC, & Harris TR (2000). Gender differences in alcohol consumption and adverse drinking consequences: Cross-cultural patterns. *Addiction*, 95(2), 251–265. [PubMed: 10723854]
- York JL, & Welte JW (1994). Gender comparisons of alcohol consumption in alcoholic and nonalcoholic populations. *Journal of Studies on Alcohol*, 55, 743–750. [PubMed: 7861804]
- Zeisser C, Thompson K, Stockwell T, Duff C, Chow C, Vallance K, ... Lucas P (2012). A 'standard joint'? The role of quantity in predicting cannabis-related problems. *Addiction Research & Theory*, 20(1), 82–92.

HIGHLIGHTS

- Optimization approaches can be used to improve clinical assessment.
- Optimal diagnostic criteria sets can be derived using complete enumeration.
- Complete enumeration can identify criteria sets common to multiple substances.

Table 1

Number of criteria sets by set size and threshold.

# criteria	# subsets by threshold										Total	
	1	2	3	4	5	6	7	8	9	10		11
1	11										11	
2	55	55									110	
3	165	165	165								495	
4	330	330	330	330							1320	
5	462	462	462	462	462						2310	
6	462	462	462	462	462	462					2772	
7	330	330	330	330	330	330	330				2310	
8	165	165	165	165	165	165	165	165			1320	
9	55	55	55	55	55	55	55	55	55		495	
10	11	11	11	11	11	11	11	11	11	11	110	
11	1	1	1	1	1	1	1	1	1	1	11	
Total	2047	2036	1981	1816	1486	1024	562	232	67	12	1	11264

Note. The set sizes range from 1 to 11 to align with the number of criteria in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).

Alcohol Use Disorder and Cannabis Use Disorder criteria endorsement rates of past year users NESARC-III ($N = 20,495$).

Table 2

Criteria	AUD ($N = 20,495$)			AUD ($N = 20,495$)			AUD ($N = 20,495$)			AUD ($N = 20,495$)			AUD ($N = 20,495$)			AUD ($N = 20,495$)			AUD ($N = 20,495$)		
	Unweighted	Weighted		Unweighted	Weighted		Unweighted	Weighted		Unweighted	Weighted		Unweighted	Weighted		Unweighted	Weighted		Unweighted	Weighted	
	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	<i>N</i>
Withdrawal	2880	14.05	18,018,738	13.14	356	9.62	2,108,011	9.41													
Failure to fulfill	550	2.68	3,166,302	2.31	86	2.32	571,833	2.55													
Social interpersonal	1703	8.31	10,761,090	7.84	271	7.32	1,694,217	7.56													
Cut down	3569	17.41	21,669,704	15.80	708	19.13	4,044,917	18.06													
Larger longer	3887	18.97	25,086,858	18.29	208	5.62	1,211,631	5.41													
Give up	530	2.59	2,929,480	2.14	143	3.86	851,471	3.80													
Physical psychological	1330	6.49	7,935,004	5.78	311	8.40	1,854,279	8.28													
Hazardous use	1931	9.42	13,207,986	9.63	850	22.97	5,511,768	24.61													
Tolerance	2031	9.91	12,415,071	9.05	554	14.97	3,472,864	15.50													
Time spent	1416	6.91	8,454,451	6.16	439	11.86	2,616,501	11.68													
Craving	2810	13.71	18,095,324	13.19	671	18.13	4,137,635	18.47													

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

Table 3

Logistic regression and degree of incremental validity adjusted for sex and age comparing diagnostic algorithms with the moderate-severe band optimal solution of AUD and/or CUD in NESARC-III.

	Any alcohol treatment use	Any cannabis treatment use	Tobacco use disorder	Any drug use disorder	Age of first drink 15 years old	Age of first cannabis use years old	Parental problem-drinking 15
DSM AUD mod +	17.95 (12.04, 26.75)	18.54 (9.24, 37.19)	3.67 (3.28, 4.11)	8.54 (6.80, 10.72)	2.56 (2.25, 2.91)	0.66 (0.58, 0.74)	2.59 (2.30, 2.92)
OS AUD	16.69 (11.18, 24.93)	12.48 (6.60, 23.62)	3.62 (3.16, 4.12)	10.05 (8.02, 12.60)	2.80 (2.47, 3.18)	0.66 (0.59, 0.73)	2.59 (2.30, 2.93)
DSM CUD mod +	3.21 (1.66, 6.21)	12.53 (6.27, 25.07)	2.74 (2.09, 3.59)	2.71 (1.85, 3.96)	1.07 (0.84, 1.38)	1.41 (1.13, 1.76)	1.21 (0.93, 1.59)
OS CUD	2.47 (1.30, 4.69)	9.63 (4.94, 18.78)	2.12 (1.69, 2.66)	2.10 (1.45, 3.06)	1.31 (1.03, 1.67)	1.61 (1.26, 2.06)	1.27 (1.01, 1.59)
DSM AUD or CUD mod +	16.58 (11.11, 24.77)	18.05 (9.09, 35.84)	3.72 (3.27, 4.17)	8.37 (6.62, 10.59)	2.47 (2.18, 2.80)	0.65 (0.58, 0.73)	2.45 (2.19, 2.75)
OS AUD OR CUD	16.53 (11.05, 24.73)	20.14 (10.73, 37.79)	3.73 (3.29, 4.24)	9.91 (7.84, 12.53)	2.80 (2.48, 3.15)	0.64 (0.58, 0.72)	2.40 (2.15, 2.68)

Added predictor	c	χ^2 diff	c	χ^2 diff	c	χ^2 diff	c	χ^2 diff	c	χ^2 diff
OS AUD	0.01	15.66*	0.01	-0.85	0.00	55.72*	0.02	0.00	0.00	80.57*
OS CUD	0.00	0.61	0.05	8.59*	0.00	5.27	0.00	0.00	0.00	9.03*
OS AUD or CUD	0.01	28.84*	0.03	19.79*	0.01	127.66*	0.02	0.00	0.00	106.23*

Note. NESARC=National Epidemiological Survey on Alcoholism and Related Conditions (Grant et al., 2003; 2015; Hasin & Grant, 2015). The top half of the table include odds ratios and confidence intervals from logistic regression models predicting the external validator comparing alternative diagnosis ONLY '0' or a diagnosis with the moderate-severe optimal solution ONLY '1', controlling for age and sex. The bottom half of this table includes results from the incremental validation where the moderate-severe optimal solution was added to logistic regression models of the alternative diagnostic rules predicting the external validator, controlling for age and sex. c=degree of incremental validity over the DSM-5 (change in concordance). The Chi-Square difference test is a one degree of freedom test. * $p < .05$. DSM-5=Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AUD=Alcohol Use Disorder; CUD=Cannabis Use Disorder; DSM-5 mod+=endorsement of at least 4 DSM-5 AUD criteria.

	Parental problem drug use	Mood disorder	Anxiety disorder	Any lifetime personality disorder	Antisocial personality disorder	Suicide attempt
DSM AUD mod +	2.09 (1.80, 2.43)	3.38 (3.04, 3.78)	2.54 (2.22, 2.91)	4.37 (3.90, 4.90)	3.30 (2.64, 4.12)	7.35 (4.59, 11.75)
OS AUD	2.14 (1.78, 2.57)	3.06 (2.68, 3.50)	2.45 (2.08, 2.90)	4.33 (3.82, 4.91)	3.81 (3.07, 4.74)	6.28 (4.15, 9.49)
DSM CUD mod +	1.68 (1.25, 2.26)	3.15 (2.44, 4.07)	2.89 (2.31, 3.61)	2.51 (1.84, 3.42)	2.17 (1.57, 3.00)	2.96 (1.57, 5.56)
OS CUD	1.64 (1.24, 2.19)	2.34 (1.83, 2.99)	2.03 (1.61, 2.55)	2.33 (1.74, 3.12)	2.35 (1.69, 3.27)	1.93 (0.98, 3.83)
DSM AUD or CUD mod +	2.17 (1.88, 2.49)	3.59 (3.25, 3.97)	2.69 (2.36, 3.08)	4.47 (4.00, 4.98)	3.68 (2.99, 4.53)	7.01 (4.49, 10.94)
OS AUD OR CUD	2.30 (1.94, 2.72)	3.23 (2.86, 3.66)	2.51 (2.14, 2.94)	4.40 (3.90, 4.97)	4.24 (3.46, 5.20)	5.45 (3.68, 8.06)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Added predictor	c	χ^2_{diff}	c	χ^2_{diff}	c	χ^2_{diff}	c	χ^2_{diff}	c	χ^2_{diff}
OS AUD	0.00	-101.11*	0.00	-16.21*	0.00	14.30*	0.01	76.73*	0.01	41.92*
OS CUD	0.00	0.44	0.00	4.66*	0.00	-5.16*	0.01	10.36*	0.01	10.35*
OS AUD or CUD	0.00	-89.86*	0.00	-5.65*	0.00	18.32*	0.01	90.13*	0.00	56.88*

Note. NESARC=National Epidemiological Survey on Alcoholism and Related Conditions (Grant et al., 2003; 2015; Hasin & Grant, 2015). The top half of the table include odds ratios and confidence intervals from logistic regression models predicting the external validator comparing an alternative diagnosis ONLY '0' or a diagnosis with the moderate-severe optimal solution ONLY '1', controlling for age and sex. The bottom half of this table includes results from the incremental validation where the moderate-severe optimal solution was added to logistic regression models of the alternative diagnostic rules predicting the external validator, controlling for age and sex. c=degree of incremental validity over the DSM-5 (change in concordance). The Chi-Square difference test is a one degree of freedom test. * $p < .05$. DSM-5=Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AUD=Alcohol Use Disorder; CUD=Cannabis Use Disorder; DSM-5 mod+=endorsement of at least 4 DSM-5 AUD criteria.

Table 4

Adjusted Rand Index (upper diagonal) and Kappa estimates (lower diagonal) across diagnostic schemes of AUD, CUD, and general SUD in NESARC-III.

	OS AUD	DSM-5 AUD	OS CUD	DSM-5 CUD	OS SUD	DSM-5 SUD
OS AUD	1	0.66	0.13	0.12	0.9	0.62
DSM-5 AUD	0.71	1	0.11	0.12	0.62	0.93
OS CUD	0.14	0.12	1	0.71	0.33	0.21
DSM-5 CUD	0.13	0.12	0.72	1	0.25	0.25
OS SUD	0.92	0.67	0.36	0.28	1	0.67
DSM-5 SUD	0.67	0.94	0.24	0.28	0.72	1

Note. NESARC = National Epidemiological Survey on Alcoholism and Related Conditions (Grant et al., 2003; 2015; Hasin & Grant, 2015). OS = Optimal Solution; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AUD = Alcohol Use Disorder; CUD = Cannabis Use Disorder; SUD = Substance Use Disorder.

Table 5

Phi coefficient (upper diagonal) and Tetrachoric (lower diagonal) correlation estimates across diagnostic schemes of AUD, CUD, and general SUD in NESARC-III.

	OS AUD	DSM-5 AUD	OS CUD	DSM-5 CUD	OS SUD	DSM-5 SUD
OS AUD	1	0.71	0.17	0.17	0.92	0.68
DSM-5 AUD	0.39	1	0.16	0.17	0.67	0.94
OS CUD	0.49	0.46	1	0.72	0.47	0.33
DSM-5 CUD	0.48	0.48	0.96	1	0.37	0.41
OS SUD	1.00	0.91	1.00	0.8	1	0.72
DSM-5 SUD	0.93	1.00	0.77	1.00	0.94	1

Note. NESARC = National Epidemiological Survey on Alcoholism and Related Conditions (Grant et al., 2003; 2015; Hasin & Grant, 2015). OS = Optimal Solution; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AUD = Alcohol Use Disorder; CUD = Cannabis Use Disorder. SUD = Substance Use Disorder.