Published in final edited form as:

Drug Alcohol Depend. 2020 November 01; 216: 108299. doi:10.1016/j.drugalcdep.2020.108299.

The Performance of DSM-5 Alcohol Use Disorder and Quantity-Frequency of Alcohol Consumption Criteria: An Item Response Theory Analysis

Tulshi D. Saha^a, S. Patricia Chou^a, Bridget F. Grant^b

^aEpidemiology and Biometry Branch, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 6700B Rockledge Drive, Room 2126, Bethesda, MD 20892

^bFedpoint Systems, LLC, Fairfax, VA 22033

Abstract

Background.—Maladaptive patterns of drinking are central to the development of AUD.

However, no DSM-5 criteria ask about patterns of alcohol use, such as 5+/4+ binge drinking. It is important to examine whether such an item would improve the diagnostic utility of the DSM-5 instrument.

Method.—Using a large representative sample of the US population, we used item response theory (IRT) methodology to examine the threshold, discrimination, and information value and differential criterion functioning of DSM-5 AUD criteria, along with a 5+/4+ drinking pattern criterion assessed at various levels of frequency.

Results.—The best fit drinking pattern criterion (defined at 5+/4+ drinking at least once a week in the past year) tapped the milder end of that continuum, which was similar to the criterion of drinking in larger amounts or for longer than intended. The new DSM-5 craving criterion was associated with mid-level values of threshold and discrimination. The AUD criteria with the addition of the 5+/4+ drinking pattern criterion demonstrated invariance across important subgroups of the population.

Conclusions.—Among the criteria with the lowest level of threshold, the drinking pattern criterion has demonstrated its utility of the DSM-5 classification by identifying clinically significant but milder AUD cases. Along with its relationship to AUD relapse, the new craving

Corresponding author Tulshi D. Saha, Tel: 301-443-0017, sahatd@mail.nih.gov.

Contributors:

Literature review: Saha, Chou, Grant

Writing first draft: Saha, Grant

Critical review of subsequent drafts: Saha, Grant, Chou

Supervision: Chou

Statistical analysis: Saha

Collected data: Saha, Chou, Grant

Conflict of Interest:

No conflicts of interest declared by any author

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

criterion tapped the moderate levels of threshold and discrimination and thus, argues for its continued inclusion in the DSM-5 AUD formulation. Study results showed that DSM-5 AUD criteria and the 5+/4+ drinking pattern criterion formed a unidimensional continuum of AUD severity.

Keywords

IRT methods; DSM-5 AUD criteria; 5+/4+ drinking pattern criterion; craving; latent construct of AUD continuum; unidimensionality

1. Introduction

In 2007, the American Psychiatric Association convened a multidisciplinary team, the Diagnostic and Statistical Manual of Mental Disorders-5 (American Psychiatric Association, 2013) Substance-Related Disorders Workgroup to identify issues in the DSM-IV (American Psychiatric Association, 1994) approach to substance use disorders and to recommend improvements for DSM-5. Two major issues identified by the workgroup were: (1) the dimensionality of the diagnostic criteria for substance use disorders; and (2) whether a new diagnostic criterion relating to quantity and frequency of substance use could be useful in identifying a substance use disorder. To address these two issues, a large number of studies used item response theory (IRT) methodology to examine the relationship between DSM-IV abuse and dependence criteria and to determine their dimensionality.

Although IRT analyses were conducted for many specific substances, by far, the most studied substance was alcohol (see Hasin et al., 2013). One major finding arising from this research was unidimensionality for DSM-IV alcohol abuse and dependence criteria except the legal problems abuse criterion, indicating that the remaining abuse and dependence criteria all indicated the same underlying construct. Among those IRT studies of DSM-IV alcohol use disorders criteria, only a few addressed adding quantity or frequency of alcohol consumption as a criterion (Beseler et al., 2010; Borges et al., 2010; Gilder et al., 2011; Hasin & Beseler, 2009; Hasin et al., 2011; McBride et al., 2011; Saha et al., 2007; Shmulewitz et al., 2010). Support for the addition of a quantity-frequency criterion (usually gender-specific quantity-frequency of drinking 5+/4+ drinks) to DSM-IV criteria was mixed with some studies finding that the criterion functioned well as a DSM-IV criterion (Beseler et al., 2010; Borges et al., 2010; Gilder et al., 2011; Saha et al., 2007) while others did not (Hasin & Beseler, 2009; McBride et al., 2011; Shmulewitz et al., 2010). Among those studies providing support for the addition of 5+/4+ drinking frequency criterion, the criterion tapped the milder end of the latent construct of AUD. These findings were important as most AUD criteria have been found to tap the more severe end of the continuum. If a drinking frequency criterion were found in this study to also tap the less severe level of the AUD continuum more mild AUD cases in need of treatment could be identified while preserving the utility of the DSM-5 classification.

All studies on IRT analyses of AUD abuse and dependence and quantity-frequency of alcohol consumption criteria were designed, in part, to inform DSM-5 classifications on issues related to the addition of a consumption variable to the classification. Thus, most of

these studies assessed the performance of the legal problem criterion, along with all other DSM-IV AUD abuse and dependence criteria, a criterion that was removed in the DSM-5 classification of AUD. Similarly, none of the IRT studies incorporating drinking quantity-frequency measures included the new DSM-5 craving criterion. Consequently, a gap in the literature was created in which IRT analyses have not been directly applied to the new DSM-5 AUD criteria nor has the performance of a quantity-frequency variable been evaluated within the context of these new criteria.

To address this gap in the literature, the objectives of this research were to use IRT methodology to: (1) determine whether DSM-5 AUD criteria along with 5+/-4+ drinking patterns criterion measured a unitary latent dimension of AUD; and (2) ascertain the presence of differential criterion functioning across important gender, age and race-ethnic subgroups of the population.

2.0 Methods

2.1. Sample

Data were derived from a large representative sample of US noninstitutionalized civilian population ($n=36,309$), 18 years or older: the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III), (Grant et al., 2014). Probability sampling was used to select respondents. Hispanics, Blacks, and Asians were oversampled. The screener- and person-level response rates were 72.0% and 84.0%, respectively, yielding a total response rate of 60.1%, comparable to most current US national surveys (Adams et al., 2013; Substance Abuse and Mental Health Services Administration, 2012). Data were adjusted for oversampling and nonresponse, then weighted to represent the US civilian population based on the 2012 American Community Survey (Bureau of the Census, 2013). Protocol and consent procedures were approved by the National Institutes of Health and Westat (data collection agency) Institutional Review Boards.

2.2. DSM-5 Alcohol Use Disorder and Quantity/Frequency Criterion

The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5) (Grant et al., 2011) was designed to measure DSM-5 AUD criteria and required the presence of at least 2 of the 11 criteria in the past 12 months preceding the interview (See Appendix). Generally, several questions operationalized each of the 11 AUD diagnostic criterion. For example, craving was defined as a positive response to either “feeling a very strong urge to drink” or “wanting a drink so badly, you couldn’t think of anything else”. Test-retest reliability of DSM-5 AUD categorical diagnoses ($\kappa = 0.60$ and $\kappa = 0.62$) and dimensional criteria scales (intraclass correlation coefficient [ICC], 0.83 and 0.85) was substantial in a large general population sample (Grant et al., 2015). Procedural validity of AUDADIS-5 and DSM-5 AUD was assessed through blind clinical reappraisal using the clinician-administered, semi-structured Psychiatric Research Interview for Substance Use and Mental Disorders, DSM-5 (PRISM-5) version (Hasin et al., 2011). The clinical reappraisal (Hasin et al., 2015) showed fair to good concordance between AUDADIS-5 and PRISM-5 AUD diagnoses ($\kappa = 0.49$ and $\kappa = 0.62$) and excellent concordance (ICC, 0.81 and 0.85) for their dimensional counterparts.

Our drinking quantity-frequency variable adopted the NIAAA definition of binge drinking (NIAAA, 2004; NIAAA, 2005), which corresponds to drinking 4 or more standard drinks (a drink equals 14 g of pure alcohol) on any day for women and 5 or more standard drinks on any day for men. Combined with binge drinking, we examined 3 levels of frequency of exceeding the daily drinking guidelines as: (1) at least once in the past year; (2) at least once a month in the past year; and (3) at least once a week in the past year. The sample for this study was restricted to 25,778 respondents classified as regular current drinkers (i.e., those who drank at least 12 drinks in the past year).

2.3 Statistical Methods

2.3.1. IRT Methodology—The choice of IRT model in this study, was the unidimensional two-parameter logistic model (Birnbaum, 1968) to define the relationship between the observed responses to the criteria and the underlying unobserved latent trait of AUD continuum:

$$p_{ij}(x_i = 1 | \theta_j) = 1 / (1 + \exp(-a_i(\theta_j - b_i)))$$

where the probability p that a person j with underlying trait θ will endorse criterion x , where b is the threshold or difficulty parameter for criterion i and a is the discrimination parameter for criterion i . (As p in the equation is the probability of endorsing a criterion, $1 - p$ is the probability of not endorsing the criterion). The a parameter measures the ability of a criterion to discriminate people who are higher versus those who are lower on the continuum. This parameter describes how strongly the criterion is related to the underlying trait or construct. The larger the a parameter (i.e. the slope at its steepest point), the greater is the discrimination of a criterion. The b parameter measures threshold of a criterion; criteria with high threshold or difficulty are endorsed less frequently and thus are more severe. The a and b parameters are plotted graphically as criterion response curves (CRCs). In these plots the b parameter represents the criterion's location along the latent continuum (located on the horizontal axis). The b parameter (threshold) is the point on the latent continuum where there is a 50% chance of the criterion to be endorsed. The a or discrimination parameter indicates how steep the slope of the CRC is at its steepest point. We used IRTPro 4.1 software (Cai, Thissen & Toit, 2011) to estimate IRT parameters, CRCs and all other IRT statistical analyses and applied Bock and Aitkin (1981) EM Algorithm (BA-EM) which is also known as marginal maximum likelihood approach. In this approach, the probability of obtaining a specific item response pattern in a population of examinees is calculated by “weighting the likelihood by the probability density of the θ vector and then integrating over the θ space” (Reckase, 2009).

The IRTPro 4.1 software also transforms the CRCs into criterion information functions (CIFs). The CIFs show where along the latent trait of AUD continuum each DSM-5 AUD criterion along with drinking quantity-frequency criterion was conveying the most information. It provides a visual representation of the information value of each criterion. Like CRCs, the latent trait of AUD continuum is plotted on the x-axis and the amount of information is plotted on the y-axis. A criterion with high discrimination (parameter ' a ') provides a higher amount of information (shows a higher peak in its CIF) and the threshold

parameter ‘b’ indicates the location where it provides the most information. Greater information is associated with greater measurement precision (Nguyen et al., 2014).

In addition to constructing CIFs for each criterion, we constructed a total CIF (TCIF) that graphically depicts the information value of the criteria collectively or in the aggregate. The TCIF curve along with its standard error of measurement curve (SEM) can be used to compare measurement precision offered by different models. The area under the TCIF curve (AUC) quantifies the total information provided by the set of criteria (Weiss & Davidson, 1981).

To be clinically meaningful, criteria should be shown to be invariant across important sociodemographic subgroups of the population. It is often unclear whether one subgroup is more likely to endorse certain AUD criteria simply because that subgroup is more likely to develop an AUD. The differential criterion function (DCF) refers to criteria that have different measurement properties for various subgroups after controlling for the overall differences between subgroups on the latent construct (Holland & Wainer, 1993). To determine whether any of the DSM-5 AUD criteria along with 5+/4+ quantity-frequency criterion displayed DCF, we compared α and λ parameters for each criterion across groups defined by gender (men as the referent category), age category (18–29 as the referent category, 30–44 and 45+ years) and race-ethnicity (White as the referent category, Black and Hispanic/others) using IRT methodology. Differences in a criterion’s discrimination parameter a between groups indicate the degree to which a criterion is related to the underlying trait differences between groups, or alternatively, that reliability of the criterion varies by group (non-uniform DCF). DCFs related to difference in a criterion’s threshold parameter b between groups (e.g. men and women) suggests that unequal levels of the trait are necessary to endorse the criterion. Wald tests were used to test the significance of DCF differences.

Criteria that demonstrate DCF need not reflect bias or variance across subgroups if the DCF occurs in opposing directions (e.g., some criteria result in greater discrimination or threshold among men while other demonstrate the opposite effect) (Cooke et al., 2001; Bolt et al., 2004). One way to evaluate the impact of identified DCF is through the construction of plots based on IRT parameter estimates, such as test expected raw score (Edelen et al., 2006). Whether criteria demonstrating significant DCF do in fact reflect invariance across subgroups can be determined if the observed DCFs cancel out at the total test (scale) score level. We plotted the expected raw scores by the latent trait of AUD continuum for age, gender and race-ethnic groups – plots referred to as the diagnostic scale function curve (DSFC). If the DSFC for groups do not substantially differ we can conclude that the significant criterion-level DCFs cancels out when considered at the total scale level. If, however, the DSFCs do differ substantially between groups, individual criteria demonstrating DCF are biased, lacking invariance across important subgroups of the population and should be eliminated

2.3.2. Assessment of IRT Model Fit—We used IRTPro 4.1 software to produce model fit statistics including Bayesian information criteria (BIC), and the M_2 family of limited information for overall goodness of fit test statistics. Tay et al., (2015) suggested using the

M_2 statistic (Maydeu-Olivares & Joe, 2005; 2006) with $p > 0.05$ and the accompanying RMSEA close to zero as good model-fit. However, the M_2 statistic follows a chi-square distribution, and like the chi-square statistics is sensitive to large sample size (as in our case). Thus, we used RMSEA, a sample-size-free index of fit computed from chi-square, instead of G square and M_2 to help assess model fit.

2.3.3. IRT Assumptions and Model Fit—The primary assumptions of the two-parameter IRT model are unidimensionality and local independence. A common approach to address the dimensionality is to use a combination of exploratory and confirmatory factor analysis (EFA and CFA) designed for ordinal data. Both EFA and CFA with model fit statistics were generated using Mplus software (Muthen & Muthen, 2010) and fitted to the data using the weighted least squares mean and variance adjusted (WLSMV) method of estimation. In EFA, we examined the unidimensionality of the data by the eigenvalues of the tetrachoric correlation matrices associated with the one-factor and two-factor models where large ratios of first to second eigenvalues indicates better fit. For CFA, we used the Hu & Bentler (1999) recommended combination of comparative fit index (CFI), Tucker-Lewis index (TLI) and Root Mean Square Error of Approximation (RMSEA), since chi-square is sensitive to the large sample sizes and may overstate the lack of fit of the structural model (Bollen, 1989). Good model fit is evidenced by a cutoff of 0.95 or above for both CFI and TLI combined with RMSEA close to 0.06.

Local independence is the assumption that, conditional on the latent variable(s), endorsement of the criteria is unrelated to one another (i.e., independent). Excess covariation among items in the residual matrix of a single-factor CFA model could indicate local dependence. Local dependence is evaluated between criterion pairs using a residual correlation matrix provided by the single factor CFA, where an absolute value in excess of 0.20 in the residual correlation indicates possible local dependence (Zhao et al, 2017).

3. Results

3.1. Prevalences and Factor Analysis

Weighted prevalences of DSM-5 AUD criteria were 1.7% for activities given up, 1.9% for neglect of roles, 4.9% for time spent, 6.3% for social/interpersonal problems, 7.1% for physical/psychological problems, 9.6% for tolerance, 10.6% for withdrawal, 10.7% for craving, 11.5% for hazardous use, 12.9% for cutdown/control, and 14.7% for larger longer (Table 1). The rates of 5+/4+ consumption criteria, exceeding daily limits at least once, at least once a month, and at least once a week during past year were 50.8%, 28.9% and 17.3%, respectively.

Fit indices indicated an excellent fit for unidimensionality to the data. The eigenvalues for the one and two factor estimations were far greater for the first factor (8.750–8.788) than for the second factor (0.551–0.583) resulting in ratios of first and second factor greater than 15 for all groups. Results from the CFA models showed excellent fit for one factor solution for the data (CFI and TLI were comparable at 0.997, and RMSEA < 0.02). Local independence was observed in all models as residual correlation for all criterion pairs was lower than the threshold (0.20). Factor loadings for all criteria were positive and significant, with

standardized values ranging from 0.74 (cutoff/control) to 0.92 (activities given up). In summary, the assumption of unidimensionality and local independence were met.

3.2. Criterion Response Curves and Criterion Information Function

Table 1 presents the IRT parameters for the two-parameter logistic models. All the three IRT models associated with different frequencies of 5+/4+ drinking criterion fitted well with RMSEAs of 0.02. The associated BIC was somewhat lower for the model with the highest frequency of the 5+/4+ drinking pattern criteria (BIC: 135851.56) compared with models with exceeding daily limits once a month or exceeding daily limits at least once a week (BIC: 141265.79 and 145263.22, respectively) indicating slightly improved fit.

The criterion response curves (CRCs) presented in Figures 1A–1C associated with each model were quite similar in terms of the relative threshold of DSM-5 AUD criteria, with the 5+/4+ drinking criterion falling along the lower level of latent AUD continuum. As expected, threshold increased as the frequency of consuming 5+/4+ drinks increased, from –0.04 for exceeding in daily limits once in the past year to 1.19 for exceeding the limits at least once a week in the past year. In all models, threshold was the greatest for the activities given up, and neglect roles (2.2) and lowest for drinking patterns criteria, larger/longer, withdrawal, cut down/control with the remaining criteria representing the intermediate threshold levels. Notably, Craving had one of the lowest threshold values. The largest (> 4.0) discrimination parameters in the sample were found for activities given up, neglect roles, time spent and physical/psychological problems criteria. Additional criteria with high discrimination parameters > 3.00 were larger/longer, social interpersonal problems and withdrawal. The criteria with lowest discrimination were hazardous use, cut down/control, tolerance and craving. The activities given up and neglect roles criteria had similar discrimination and threshold parameters with overlapping CRCs.

The criterion information function curves (CIFs) presented in Figures 2A–2C for all AUD criteria and each 5+/4+ drinking criteria contributed different amounts of information and all (except drinking criteria) were contributing information with similar ranges at the higher end of the latent trait. All models indicated that four criteria (time spent, neglect roles, activities given up and physical/psychological problems) had very high discrimination parameters (i.e., $a > 4.00$), which resulted in high information values for these criteria. The larger/longer criteria demonstrated the greatest information values across a boarder range of the latent AUD continuum; however, less information was provided by the 5+/4+ drinking criterion. The criteria that associated with the least information (CIF) were cut down/control and hazardous use.

The TCIF is presented in Figure 3. For all models, the continuum provided most information for individuals with moderate to severe level of the latent AUD continuum but less information for individuals in the low level of the continuum. The SEM curves clearly show less error in measurement at the mid- and higher end of the latent trait compared with lower end of the latent AUD continuum. The TCIFs for the three models associated with frequency of the 5+/4+ drinking criterion were identical and the area under the TCIF curve (AUC) was also similar for all three models and did not significantly differ. [The AUC for the model with exceeding daily limits at least once: 44.24; 95% CI (37.31–51.161), AUC for model

with exceeding daily limits at least once a month: 43.79; 95% CI (36.51–51.07) and AUC for model with exceeding daily limits at least once a week: 43.46; 95% CI (36.03–50.88)]. As a point of comparison, the AUC for the model including only the 11 DSM-5 AUD diagnostic criteria was 41.80 indicating less information than when the model included a drinking frequency measure.

3.3. Differential Criterion Function (DCF)

An advantage of IRT modeling is that it provides an elegant framework for the evaluation of between-group differences in criterion functioning (Relse & Rodriguez, 2016). AUD criteria such as larger/longer, cut down, craving, hazardous use, tolerance, withdrawal and drinking criteria showed significant DCF by gender, age group and race-ethnicity group. However, more criteria showed DCF between 18–29 and 30–44-year-olds and the older age group than between age groups 30–44 and 45+. Similarly, a larger number of criteria produced DCF between White and non-White group as compared to DCF between Blacks and Hispanics/others. Women had a lower standardized latent trait mean and higher standard deviation than men. The latent trait standardized mean was lower and standard deviation was higher for the older age groups (age 30 and older) as compared to younger age group (age 18–29), and the latent trait mean was lower and standard deviation higher or the same as non-Whites than Whites.

The diagnostic scale function curves (DSFC) were nearly identical, implying that impact of identified DCF on the estimated population subgroup differences in diagnostic scale/expected scores of the total 12 criteria (11 DSM-IV AUD criteria and 5+/4+ drinking criteria) is negligible (Figures 4A–4C). These curves were also identical for each of the three models associated with frequency of the 5+/4+ drinking criterion. For illustrative purposes we present only the model that included exceeding the 5+/4+ drinking criterion at least once a week in the past year.

4.0 Discussion

Similar to the majority of studies that examined DSM-IV abuse and dependence criteria along with a 5+/4+ drinking variable (Beseler et al., 2009; Borges et al., 2010; Gilder et al., 2011; McBride et al., 2011; Saha et al., 2007; Shumulewitz et al., 2010), DSM-5 AUD criteria and the 5+/4+ drinking pattern criterion, assessed at three levels of frequency, were arrayed along a continuum of difficulty. Model fit was best when exceeding the daily drinking limits at least once a week during the past year was included in the IRT model. Although we observed differential criterion functioning by gender, age and race-ethnicity for certain criteria, analyses at the diagnostic scale level overall indicated that the impact of DCF on population subgroup differences was negligible. Taken together those results support the reliability and validity of the DSM-5 AUD diagnosis (including the 5+/4+ drinking pattern criterion) and the invariance of its associated criteria across important subgroups of the population.

Consistent with results of earlier studies (Beseler et al., 2010; Borges et al., 2010, Gilda et al., 2011; Saha et al., 2007; Shumulewitz et al., 2010), the larger/longer criterion was one of the least severe criteria with good discrimination. Because of these properties, Saha and his

colleagues (2007) suggested that the larger/longer criterion may be a bridging criterion that links the less severe with the more severe end of the AUD continuum. They reasoned that since the larger/longer criterion represents a high-risk drinking pattern, it is possible that other high-risk drinking patterns that incur risk of AUD, e.g. drinking that exceeds the nationally recommended guidelines as measured here, may be good candidates to represent the lower range of the AUD continuum. This study and others (Beseler et al., 2010; Borges et al., 2010; Gilders et al., 2011; Saha et al., 2007; Shumulewitz et al., 2010) found that the 5+/4+ drinking pattern criterion had among the lowest threshold (with intermediate discrimination) levels relative to DSM-IV abuse and dependence criteria. These results suggest that the addition of the 5+/4+ drinking pattern criterion would preserve the utility of the DSM-5 classification while helping to identify clinically significant but more mild cases of AUD, whom may be in need of treatment.

The relative ordering of the difficulty of DSM-5 AUD criteria were largely similar to prior studies conducted with the inclusion of the 5+/4+ drinking pattern variable (Beseler et al., 2010; Borges et al., 2010; Gilders et al., 2011; Saha et al., 2007; Shumulewitz et al., 2010) and those examining DSM-IV abuse and dependence criteria alone (Langenbucher et al., 2004; Proudfoot et al., 2006). The time spent, neglect of roles, activities given up and physical/psychological problems criteria were among the most severe (and informative) criteria, the withdrawal, cut down/control and tolerance criteria were among the lowest difficulty indicators with the remaining criteria overlapping in the middle range of the observed difficulty continuum. Taken together, these findings support the DSM-5 workgroup's decision to combine the majority of DSM-IV abuse and dependence criteria into a single diagnosis of AUD.

Craving was a new criterion added to the DSM-5 classification of AUD. In our study, we found that craving was associated with moderate difficulty and discrimination, a result consistent with prior research (Casey et al., 2012; Cherpital et al., 2010; Keys et al., 2011; Mewton et al., 2011). Our finding showed that craving tapped into the same underlying latent construct, the DSM-5 AUD continuum. However, the similar parameterization of craving with several other criteria in the mid-range of difficulty and discrimination may reflect psychometric redundancy. However, the important relationship between craving and AUD relapse (Bottlender & Soyka, 2004; Evren et al., 2010) argues for its retention as an DSM-5 AUD criterion.

Limitations are noted. Although self-report measures are subject to recall bias, we limited the present analyses to past 12-month criteria to help mitigate this bias. We acknowledge that the results of this study may not be generalized to other cultures. Craving was operationalized with 2 question items. However, craving is an extremely complex process, encompassing phenomenological, environmental, cognitive and affective states. Research on the development of craving measure encompassing these domains is warranted.

In summary, this study found that DSM-5 AUD criteria (along with the 5+/4+ drinking pattern criterion) were arrayed along a unidimensional continuum of disorder spectrum supporting the DSM-5 revision that combined most DSM-IV abuse and dependence criteria into a single diagnosis. The use of threshold, discrimination and/or information values

associated with each criterion can be used as weights in the development of dimensional scales, with the most difficult or discriminating criterion given greater weight. The development of valid and reliable AUD scales holds great promise in genetic and neuroscience alcohol research that has heretofore been hindered by categorical AUD representations.

Representing the lower range of the AUD continuum, the 5+/4+ drinking pattern criterion was considered a good candidate because of its utility in identifying clinically significant but milder cases of AUD. Although the new DSM-5 craving criterion tapped the mid-range of the difficulty continuum, along with other AUD criteria, its important relationship to AUD relapse argues for its retention in the DSM-5 AUD conceptualization. Further research on the utility of DSM-5 AUD criteria (and the 5+/4+ drinking pattern criterion) is warranted. Research using IRT methodology should also focus on examining additional criterion that could improve the reliability, validity and cross-national applicability of the DSM-5 AUD classification.

Acknowledgments

Role of Funding Source:

The NESARC-III was sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA), with supplemental support from the National Institute on Drug Abuse. Support is also acknowledged from the Intramural program, NIAAA, NIH. Sponsors and funders of the NESARC-III had no role in the design and conducted of the study; collection, management analysis, and interrelation of data; and preparation, review and approval of the manuscript.

Appendix:

The DSM-5, alcohol use disorder criteria included (1) drinking larger amounts or for longer periods than intended; (2) persistent desire or unsuccessful efforts to cut down or control drinking; (3) a great deal of time spent in activities to obtain alcohol to drink, or to recover from its effects; (4) craving; (5) failure to fulfill major role obligations at work/school/home; (6) social or interpersonal problems; (7) giving up or reducing important activities in favor of drinking; (8) use in hazardous situations; (9) continued drinking despite knowledge of a physical or psychological problem caused or exacerbated by drinking; (10) tolerance; (11) withdrawal symptoms or withdrawal relief/avoidance.

References

- Adams PF, Kirzinger WK, Martinez M, 2013. Summary health statistics for the U.S. population: National Health Interview Survey, 2012. *Vital Health Stat* 10 259, 1–95.
- American Psychiatric Association, 1992. *Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition*. Arlington, VA: American Psychiatric Association.
- Beseler CL, Taylor LA, Leeman RF, 2010. An item-response theory analysis of DSM-IV alcohol-use disorder criteria and “binge” drinking in undergraduates. *J. Stud. Alcohol Drugs* 71, 418–423. [PubMed: 20409436]
- Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* 46, 443–459 (1981). 10.1007/BF02293801

- Bolt DM, Hare RD, Vitale JE, Newman JP. 2004. A multigroup item response theory analysis of the Psychopathy Checklist-revised. *Psychol Assess.*16:155–168 [PubMed: 15222812]
- Borges G, Ye Yu, Bond J, Cherpitel CJ, Cremonte M, Moskalewicz J, Swiatkiewicz G, Rubio-Stipec M, 2010. The dimensionality of alcohol use disorders and alcohol consumption in a cross-national perspective. *Addiction* 105, 240–254. [PubMed: 20078482]
- Bottlender M, Soyka M, 2004. Impact of craving on alcohol relapse during, and 12 months following outpatient treatment. *Alcohol Alcohol*, 39, 357–361. [PubMed: 15208171]
- Bureau of the Census, 2013. American Community Survey, 2012. Suitland, MD: Bureau of the Census.
- Cai L, Thissen D, & du Toit SHC (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International
- Casey M, Adamson G, Shevlin M, McKinney A, 2012. The role of craving in AUDs: dimensionality and differential functioning in the DSM-5. *Drug Alcohol Depend.* 125, 75–80. [PubMed: 22516145]
- Centers for Disease Control and Prevention, 2013. Unweighted response rates for NHANES 2011–2012. Atlanta, GA: Centers for Disease Control and Prevention.
- Cherpitel CJ, Borges G, Ye Yu, Bond J, Cremonte M, Moskalewicz J, Swiatkiewicz G, 2010. Performance of a craving criterion in DSM alcohol use disorders. *J. Stud. Alcohol Drugs*, 71, 674–684. [PubMed: 20731972]
- Cooke DJ, Kosser DS, Michie C. 2001. Psychopathy and ethnicity: structural, item and test generalizability of the Psychopathy Checklist – Revised (PCL-R) in Caucasian and African American participants. *Psychol Assess.*13:531–542. [PubMed: 11793896]
- Evren C, Cetin R, Durkaya M, Dalbudak E, 2010. Clinical factors associated with relapse in male alcohol dependents during six-month follow-up. *Behav. Cogn. Psychother* 20, 14–22.
- Gilder DA, Gizer IR, Ehlers CL, 2011. Item response theory analysis of binge drinking and its relationship to lifetime alcohol use disorder symptom severity in an American Indian Community Sample. *Alcohol Clin. Exp. Res* 35(5), 984–995. [PubMed: 21314696]
- Grant BF, Goldstein RB, Chou SP, et al., 2011. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition Version (AUDADIS-5). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- Grant BF, Amsbary M, Chu A, et al., 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 BF, Goldstein RB, Smith SM, et al., 2015. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): reliability of substance use and psychiatric disorder modules in a general population sample. *Drug Alcohol Depend.* 148(1), 27–33. [PubMed: 25595052]
- Hasin DS, Beseler CL, 2009. Dimensionality of lifetime alcohol abuse, dependence and binge drinking. *Drug Alcohol Depend.* 101, 53–61. [PubMed: 19095379]
- Hasin DS, Aivadyan C, Greenstein E, Grant BF, 2011. Psychiatric Research Interview for Substance Use and Mental Disorders, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (PRISM-5) Version. New York, NY: Columbia University, Department of Psychiatry.
- Hasin DS, O'Brien CP, Auriacombe M. et al., 2013. DSM-5 criteria for substance use disorders: recommendation and rationale. *Am. J. Psychiatry* 170(8), 834–851. [PubMed: 23903334]
- Hasin DS, Greenstein E, Aivadyan C, et al., 2015. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): procedural validity of substance use disorders modules through clinical re-appraisal in the general population sample. *Drug Alcohol Depend.* 148(1), 40–46. [PubMed: 25604321]
- Holland PW, & Wainer H. (Eds). 1993. Differential item functioning. Lawrence Erlbaum Associates, Inc
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. 1999. *Struct Equat Model.*26:1–55.
- Keyes KM, Krueger RF, Grant BF, Hasin DS, 2011. Alcohol craving and the dimensionality of alcohol disorders. *Psychol. Med* 41, 629–640. [PubMed: 20459881]

- Langenbucher JW, Labouvie E, Martin CS, Sanjuan PM, Bavly L, Kirisci L. 2004. An application of item response theory analysis to alcohol, cannabis and cocaine criteria in DSM-IV. *J Abnorm Psychol.* 13:72–80
- McBride O, Teesson M, Baillie A, Slade T, 2011. Assessing the dimensionality of lifetime DSM-IV alcohol use disorders and a quantity-frequency alcohol use criterion in the Australian population: a factor mixture modelling approach. *Alcohol.* 46(3), 333–341.
- Maydeu-Olivares A, Joe H. 2005. Limited and full information estimation and goodness-of-fit testing in 2n contingency tables: a unified framework. *J. Am. Stat. Assoc* 100:1009–20
- Mewton L, Slade T, McBride O, Grove R, Teesson M, 2011. An evaluation of the proposed DSM-5 alcohol use disorder criteria using Australian national data. *Addict.* 106, 941–950.
- Muthén BO, Muthén LK. *Mplus: Statistical Analysis with Latent Variables (Version 6.1)* Muthén & Muthén, Inc; Los Angeles, CA: 2010.
- Nguyen TH, Han HR, Kim MT. et al. An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *Patient* 7: 23. 10.1007/s40271-013-0041-0.
- Proudfoot H, Baillie AJ, Teesson M. 2006. The structure of alcohol dependence in the community. *Drug Alcohol Depend.* 81:21–26 [PubMed: 16005578]
- Reckase MD 2009. *Multidimensional Item response Theory*. New York, NY. Springer
- Reise S, & Rodriguez A. 2016. Item response theory and the measurement of psychiatric constructs: Some empirical and conceptual issues and challenges. *Psychological Medicine*, 46(10), 2025–2039. doi:10.1017/S0033291716000520 [PubMed: 27056796]
- Saha TD, Stinson FS, Grant BF, 2007. The role of alcohol consumption in future classifications of alcohol use disorders. *Drug Alcohol Depend.* 89, 82–92. [PubMed: 17240085]
- Shmulewitz D, Keyes K, Beseler C, Aharonovich E, Aivadyan C, Spivak B, Hasin D, 2010. The dimensionality of alcohol use disorders: results from Israel. *Drug Alcohol Depend.* 111, 146–154. [PubMed: 20537809]
- Abuse Substance and Mental Health Services Administration, 2012. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings, Appendix B: Statistical Methods and Measurement. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Tay L, Meade A W, Cao M. 2015 An Overview and Practical Guide to IRT Measurement Equivalence Analysis. 2015. *Organizational Research Methods*. Vol 28. Issue 1.
- United States Department of Agriculture, 2014. *Dietary Guidelines-Alcohol*. Washington, DC: U.S. Department of Agriculture.
- United States Department of Health and Human Services, 2014. 2015–2020 Dietary Guidelines. Washington, DC: U.S. Department of Health and Human Services.
- Weiss D and Davidson M L. 1981. Test Theory and Methods. *Ann Rev Psychol.* 32: 629–58.
- Zhao Y, Chan W, Lo BCY. 2017. Comparing five depression measures in depressed Chinese patients using item response theory: an examination of item properties, measurement precision and score comparability. *Health and Quality of Life Outcomes.* 15:60 [PubMed: 28372559]

Highlights

- DSM-5 AUD criteria and 5+/4+ drinking pattern criteria all formed latent construct of AUD continuum.
- The 5+/4+ drinking pattern criterion tapped the milder end of latent AUD severity continuum.
- The drinking pattern criterion once a week in past year had the best model fit.
- Craving criterion were associated with mid-level values of threshold and discrimination.
- Most severe AUD criteria were time spent, activities given up, neglect of role and psychological/physical problems.

Fig 1A:

Criteria response curves for DSM-5 AUD criteria and 5+/4+ drinking at least once (occasionally) in the past year

Fig 1B:

Criteria response curves for DSM-5 AUD criteria and 5+/4+ drinking at least once a month in the past year

Fig 1C:

Criteria response curves for DSM-5 AUD criteria and 5+/4+ drinking at least once a week in the past year

Fig 2A:

Criteria information curves for DSM-5 AUD criteria and 5+/4+ drinking at least once in the past year

Fig 2B:

Criteria information curves for DSM-5 AUD criteria and 5+/4+ drinking at least once a month in the past year

Fig 2C:

Criteria information curves for DSM-5 AUD criteria and 5+/4+ drinking at least once a week in the past year

Area Under Curve (AUC):

	Model with at least once (occasional)	Model with at least once a month	Model with at least once a week
Value	44.24601	43.78968	43.45796
LB	37.31262	36.50963	36.03129
UB	51.15939	51.06973	50.88463

Figure 3:
Total criterion information function and standard error (SE) of measurement.

Figure 4:
Diagnostic Scale Function Curve (DSFC) for sex (A), age (B) and race-ethnicity (C) for DSM-5 AUD criteria and drinking 5+/4+ at least once a week in the past year

Table 1:

Prevalence, Factor Loadings and Criterion Response Parameters: Alcohol Use Disorder (AUD) with Frequency of 5+/4+ Drinking Criterion

Past Year DSM-5 Alcohol Use Disorder (AUD) Criteria and Level of Intense Drinking Criteria	Weighted Prevalence (%)	Factor loadings of AUD Criteria with exceeded daily limits of drinking			Two Parameter IRT Model		Two Parameter IRT Model		Two Parameter IRT Model	
					Model with exceeded daily limits at least once		Model with exceeded daily limits at least once a month		Model with exceeded daily limits at least once a week	
		At least once (EX1)	At least once a month (EXMO)	At least once a week (EXWK)	Discrimination a (SE)	Threshold b (SE)	Discrimination a (SE)	Threshold b (SE)	Discrimination a (SE)	Threshold b (SE)
Larger/longer (LL)	14.7	0.870	0.863	0.860	3.38 (0.07)	1.17 (0.01)	3.28 (0.07)	1.18 (0.01)	3.23 (0.07)	1.18 (0.01)
Cutdown/control (CC)	12.9	0.744	0.748	0.749	2.06 (0.04)	1.42 (0.02)	2.09 (0.04)	1.41 (0.02)	2.09 (0.04)	1.41 (0.02)
Time spent (TS)	4.9	0.907	0.910	0.908	4.18 (0.12)	1.73 (0.02)	4.22 (0.13)	1.73 (0.02)	4.18 (0.12)	1.74 (0.02)
Craving (CR)	10.7	0.827	0.827	0.828	2.77 (0.06)	1.46 (0.02)	2.75 (0.06)	1.46 (0.02)	2.76 (0.06)	1.46 (0.02)
Neglect roles (NR)	1.9	0.910	0.908	0.909	4.53 (0.18)	2.15 (0.02)	4.5 (0.18)	2.15 (0.02)	4.47 (0.18)	2.16 (0.02)
Social/interpersonal problems (SI)	6.3	0.846	0.849	0.848	3.18 (0.08)	1.72 (0.02)	3.2 (0.08)	1.72 (0.02)	3.17 (0.08)	1.73 (0.02)
Activities given up (AG)	1.7	0.921	0.921	0.925	4.59 (0.19)	2.16 (0.02)	4.58 (0.19)	2.16 (0.02)	4.59 (0.19)	2.17 (0.02)
Hazardous use (HU)	11.5	0.737	0.739	0.737	2.05 (0.04)	1.59 (0.02)	2.05 (0.04)	1.59 (0.02)	2.03 (0.04)	1.6 (0.02)
Physical / psychological problems (PP)	7.1	0.899	0.900	0.904	4.06 (0.11)	1.57 (0.01)	4.08 (0.11)	1.57 (0.01)	4.1 (0.11)	1.57 (0.01)
Tolerance (TL)	9.6	0.787	0.791	0.791	2.46 (0.05)	1.56 (0.02)	2.48 (0.05)	1.56 (0.02)	2.47 (0.05)	1.56 (0.02)
Withdrawal (WD)	10.6	0.866	0.860	0.859	3.23 (0.07)	1.39 (0.01)	3.14 (0.07)	1.4 (0.01)	3.11 (0.07)	1.4 (0.01)
Exceeded daily limits at least once (EX1)	50.8	0.808			2.34 (0.05)	-0.04 (0.01)				
Exceeded daily limits at least once a month (EXMO)	28.9		0.773				2.04 (0.04)	0.67 (0.01)		
Exceeded daily limits at least once a week (EXWK)	17.3			0.762					1.94 (0.04)	1.19 (0.02)
Ratio of Eigen values 1 and 2		15.07	15.41	15.88						
CFI		0.997	0.997	0.997						

Past Year DSM-5 Alcohol Use Disorder (AUD) Criteria and Level of Intense Drinking Criteria	Weight ed Prevalence (%)	Factor loadings of AUD Criteria with exceeded daily limits of drinking			Two Parameter IRT Model		Two Parameter IRT Model		Two Parameter IRT Model	
					Model with exceeded daily limits at least once		Model with exceeded daily limits at least once a month		Model with exceeded daily limits at least once a week	
		At least once (EX1)	At least once a month (EXMO)	At least once a week (EXWK)	Discrimination a (SE)	Threshold b (SE)	Discrimination a (SE)	Threshold b (SE)	Discrimination a (SE)	Threshold b (SE)
TLI		0.997	0.997	0.997						
Root mean square residual (RMSE) (95% CI)		0.014 (0.013– 0.016)	0.015 (0.013– 0.016)	0.014 (0.013– 0.016)						
Bayesian Information Criterion (BIC) [Smaller is better and marked as ✓]					145263.22		141265.79		135851.56 ✓ Best model	