



Published in final edited form as:

*Drug Alcohol Depend.* 2010 February 1; 107(1): 31–38. doi:10.1016/j.drugalcdep.2009.08.019.

## A MULTIDIMENSIONAL ASSESSMENT OF THE VALIDITY AND UTILITY OF ALCOHOL USE DISORDER SEVERITY AS DETERMINED BY ITEM RESPONSE THEORY MODELS

Deborah A. Dawson<sup>1</sup>, Tulshi D. Saha<sup>1</sup>, and Bridget F. Grant<sup>1</sup>

Laboratory of Epidemiology and Biometry, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda MD, USA

### Abstract

**Background**—The relative severity of the 11 DSM-IV alcohol use disorder (AUD) criteria are represented by their severity threshold scores, an item response theory (IRT) model parameter inversely proportional to their prevalence. These scores can be used to create a continuous severity measure comprising the total number of criteria endorsed, each weighted by its relative severity.

**Methods**—This paper assesses the validity of the severity ranking of the 11 criteria and the overall severity score with respect to known AUD correlates, including alcohol consumption, psychological functioning, family history, antisociality, and early initiation of drinking, in a representative population sample of U.S. past-year drinkers (n=26,946).

**Results**—The unadjusted mean values for all validating measures increased steadily with the severity threshold score, except that legal problems, the criterion with the highest score, was associated with lower values than expected. After adjusting for the total number of criteria endorsed, this direct relationship was no longer evident. The overall severity score was no more highly correlated with the validating measures than a simple count of criteria endorsed, nor did the two measures yield different risk curves. This reflects both within-criterion variation in severity and the fact that the number of criteria endorsed and their severity are so highly correlated that severity is essentially redundant.

**Conclusions**—Attempts to formulate a scalar measure of AUD will do as well by relying on simple counts of criteria or symptom items as by using scales weighted by IRT measures of severity.

### Keywords

Item response theory; AUD severity; validity; alcohol use disorder; severity; validation

---

Please send all correspondence to: Deborah A. Dawson, Ph.D., NIAAA/LEB, 5635 Fishers Lane, Room 3071, Bethesda, MD 20892-9304 Telephone: 301-435-2244 Fax: 301-443-1400 ddawson@mail.nih.gov.

**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 citable 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.

*Disclaimer:* The views and opinions expressed in this paper are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies or the U.S. government.

## 1.0 INTRODUCTION

Since the conceptualization of the alcohol dependence syndrome (Edwards and Gross, 1976), in which alcohol problems were characterized as lying along a continuum of severity, clinicians and researchers have recognized that continuous scales of alcohol problem severity offer more information than dichotomous diagnostic measures such as alcohol abuse, harmful use, and alcohol dependence. At the lower end of their values, severity scales help to distinguish risk levels among individuals who do not yet meet the clinical criteria for an alcohol use disorder (AUD), identifying those for whom some sort of brief intervention may forestall the development of an AUD. At the upper end of their values, severity scales may help to determine the need for specific types of treatment, and they also offer a metric with which treatment progress and/or AUD progression can be monitored.

Continuous measures of alcohol problem severity include many widely-used scales, such as the Severity of Alcohol Dependence Questionnaire (SADQ) (Stockwell et al., 1979), the Alcohol Dependence Scale (ADS) (Skinner and Allen, 1982), the Addiction Dependence Data and Short Alcohol Dependence Data scales (Raistrick et al., 1983), the Addiction Severity Index (ASI) (McLellan et al., 1980), the Alcohol Use Disorder Identification Test (AUDIT) (Saunders et al., 1993), and the Leeds Dependence Questionnaire (Raistrick et al., 1994). The reliability and internal consistency of these scales have been documented extensively in clinical and general population samples. Their concurrent and convergent validity have been established in comparisons with other severity scales and/or alcohol dependence screeners (Butler et al., 2009; Daeppen et al., 1995; Thomas and McCambridge., 2008), alcohol consumption patterns and biomarkers (Donovan et al., 2006; Read et al., 2007; Thomas and McCambridge, 2008; Wood et al., 2003), psychological functioning and social adjustment (Butler et al., 2009; Lima et al., 2005), and treatment goals and progress (Donovan et al., 2006).

A number of recent studies have employed item response theory (IRT) models, sometimes called latent trait models, to study the severity of AUD. After confirming the unidimensionality of a set of alcohol problem items, e.g., AUD symptoms or criteria, these models estimate parameters that reflect the relative severity of each item along a continuum of AUD severity. Two-parameter models also estimate how effectively each item discriminates individuals at different levels of severity. A recent study using data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) established that the 11 DSM-IV (American Psychiatric Association, 1994) abuse and dependence criteria were interspersed in terms of severity. Some of the abuse criteria (e.g., interpersonal problems caused by drinking) lay at the highest end of the severity scale and some of the dependence criteria (e.g., tolerance and withdrawal) lay at the lowest end, thus calling into question the DSM-IV hierarchy of abuse and dependence (Saha et al., 2006). Likewise, Proudfoot et al. (2006) found that the abuse criteria of hazardous drinking legal problems and drinking despite social/interpersonal problems were among the most severe criteria in an IRT analysis of a representative Australian population sample. Kahler and Strong (2006) reported similar findings when modeling the NESARC abuse and dependence symptoms rather than criteria, and they also found that any given criterion was associated with varying levels of severity depending on how it was operationalized. The impact of operationalization on criterion severity was also highlighted by a recent report (Harford et al., 2009) based on the National Survey on Drug Use in Households (NSDUH), in which the ranking of the 11 AUD criteria not only varied by age and gender but also differed from the ranking based on the NESARC (Saha et al., 2006).

Given recent arguments for and evidence of the desirable qualities of dimensional as opposed to categorical measures of AUD (Hasin and Beseler, 2009; Hasin et al., 2006; Helzer et al., 2006, 2007), IRT models also have been used to create summary measures of AUD severity.

Kahler et al. (2003) used IRT models to derive a new, empirically-based scoring system for selected ADS items. The reduced ADS was as highly correlated as the full ADS with frequency of heavy drinking, the Alcohol Use Disorder Identification Test (AUDIT, Saunders et al., 1993) and the drug and alcohol subscales of the Psychiatric Diagnostic Screening Questionnaire (Zimmerman and Mattia, 2001) in a sample of high-risk drinkers. Using data from the National Longitudinal Study of Adolescent Health, Kahler et al. (2009) reported that consumption and problem measures could be combined to create a valid index of alcohol involvement. Items ranged from having consumed 12+ drinks in the past year or gotten drunk at the low end of the severity spectrum to having school problems and daily drinking at the high end.

To date, the IRT-derived severity rankings of the DSM-IV abuse and criteria as measured in the NESARC have not been validated against other indicators of AUD severity; nor have the IRT severity parameters themselves been employed in the creation of severity-weighted AUD scale. Accordingly, this analysis has two objectives. First, it examines the severity ranking of the criteria with multiple correlates of AUD, including consumption, psychological functioning, family history, antisociality and early initiation of drinking, to see if individuals endorsing successively more severe criteria have correspondingly higher values for the correlates (lower in the case of psychological functioning). This is done both before and after controlling for other criteria endorsed, in order to determine whether some criteria are inherently more severe than others or whether apparent differences in severity merely reflect a greater number of positive criteria, i.e., a broader and therefore more severe range of alcohol problems, among individuals endorsing the more severe criteria. Second, it tests whether severity-weighted counts of endorsed criteria are more strongly associated with these AUD correlates than simple unweighted counts of endorsed criteria, in order to see whether the creation of a weighted scale is worth the time and effort in terms of yielding a more sensitive prognostic tool. Additionally, it examines the implications, both for the severity ranking of criteria and their use in creating a continuous AUD measure, of the specific symptom items endorsed, i.e., the operationalization of each criterion. The analysis is based on a nationally representative sample of U.S. past-year drinkers 18 years of age and older, using data from the Wave 1 NESARC.

## 2.0 METHODS

### 2.1 Sample

The data used in this analysis came from Wave 1 of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), designed by the National Institute on Alcohol Abuse and Alcoholism (Grant et al, 2003a). The 2001–2002 Wave 1 NESARC sample represented U.S. adults 18 years or older residing in households and noninstitutional group quarters in all 50 states and the District of Columbia ( $n = 43,093$ , response rate = 81.0%). Data were weighted to reflect design characteristics, oversampling of Blacks, Hispanics and young adults, and nonresponse relative to sociodemographic characteristics. The weighted data were then adjusted to match the civilian, noninstitutionalized population of the United States on socioeconomic variables based on the 2000 Decennial Census. Data were collected in personal interviews. All potential respondents were informed in writing about the nature of the survey, the statistical uses of the survey data, the voluntary aspect of their participation and the Federal laws that rigorously provide for the confidentiality of identifiable survey information. Only respondents consenting to participate after receiving this information were interviewed. The research protocol, including informed consent procedures, received full ethical review and approval from the U.S. Census Bureau and the U.S. Office of Management and Budget. This analysis is based on a subsample of respondents who had consumed at least one drink in the year preceding the Wave 1 interview ( $n=26,946$ ).

## 2.2 Measures of Past-Year AUD Severity

Past-year AUD was assessed by means of 33 symptom item indicators in the Alcohol Use Disorder and Associated Disabilities Interview Schedule – DSM-IV Version (AUDADIS-IV), a structured diagnostic instrument designed for use by lay interviewers (Grant et al., 2001). These symptom items reflected the four DSM-IV alcohol abuse criteria (recurrent failure to fulfill role obligations because of alcohol use, recurrent use in hazardous situations, recurrent alcohol-related legal problems, continued use despite awareness of recurrent interpersonal problems) and the seven DSM-IV alcohol dependence criteria (tolerance, withdrawal or relief/avoidance of withdrawal, use in larger amounts or for longer than intended, persistent desire/unsuccessful attempts to control use, a great deal of time spent in alcohol-related activities, important activities given up because of alcohol, continued use despite knowledge of physical or psychological problems). A simple count of the number of criteria positively endorsed (range = 0 to 11) was calculated for each respondent. Test-retest reliability was good both for the dichotomous measures of any past-year alcohol abuse or dependence ( $\kappa = 0.74$ ) and for the scalar measures of numbers of abuse and dependence symptoms (intraclass correlation coefficients of 0.76 and 0.78, respectively)(Grant et al., 2003b).

Severity threshold parameters representing the relative severity of each AUD criterion along a continuum of the latent trait of alcohol use disorder, with larger parameters reflecting greater severity, were estimated from a two-parameter logistic model using item response theory (IRT) models (see Section 2.4, Analysis). These parameters are directly proportional to the frequency with which the criteria were endorsed and correspond to the  $b$  parameters from the model. (The  $a$  parameters, which measure the ability of the criteria to discriminate people who are higher and lower on the continuum of severity, were not used in this analysis.) A weighted severity score (range = 0 to 26.07) was calculated by summing the values of the severity threshold parameters across all positively endorsed criteria.

## 2.3 Validating Measures

For each of four beverage types (coolers, beer, wine and spirits) and for all types of alcohol combined, the NESARC collected information on overall frequency of drinking, usual quantity consumed, largest quantity, and frequencies of drinking the largest quantity and of drinking 5 + drinks. Average daily volume (ADV) of ethanol intake (in ounces) was based on a formula that incorporated all of the quantity data (usual, largest and 5+), weighted by their respective frequencies, and was set to the largest of the sum of the four beverage-specific volumes or the volume based on all types of alcohol combined. To reduce skew and avoid undue influence of outliers, ADV was top-coded to a high of 32 oz. (affecting less than 0.1% of cases). The measures of overall frequency of drinking, largest quantity of drinks and frequency of drinking 5+ drinks used in this analysis were taken from the series of questions for all types of alcohol combined. Frequencies refer to days per year. All consumption measures refer to the 12 months preceding the Wave 1 interview. Test-retest reliability was good for average daily volume of ethanol intake and frequency of risk drinking, with intraclass correlation coefficients (ICC) of .72 and .70, respectively and excellent for maximum quantity of drinks, ICC = .84 (Grant et al., 2003b).

Psychological functioning was measured using the norm-based mental component score (NBMCS) from Version 2 of the Short Form 12 Health Survey (SF-12v2), standardized to a mean of 50 (range = 0 to 100) using standard norm-based scoring techniques (Ware et al., 2002), with higher scores indicative of better functioning. Alternate-form reliabilities of these scales are excellent in the general U.S. population, 0.89 and 0.86, respectively (Ware et al., 2002). The SF-12 scales have high levels of test-retest reliability (Bohannon et al., 2004), internal consistency (King et al., 2005; Resnick and Nahm, 2001) and convergent validity with respect to other indexes of physical and mental health (Salyers et al., 2000). Family history

density of alcoholism was constructed using the algorithm proposed by Stoltenberg et al. (1998), where each alcoholic parent contributes 0.5 and each alcoholic grandparent contributes 0.25 (range = 0 to 2). Number of lifetime antisocial personality disorder (ASPD) symptoms (range = 0 to 33) was calculated from a list of 33 symptom items used to operationalize conduct disorder and ASPD in the AUDADIS-IV. Initiation of drinking before age 15 was determined by self report to a question asking when the respondent first started drinking, not counting small tastes or sips.

## 2.4 Analysis

Given strong evidence of the unidimensionality of AUD from prior studies (Harford et al., 2009; Kahler et al., 2006, 2009; Proudfoot et al., 2006; Saha et al., 2006, 2007), the fit of a single-factor model was tested by means of exploratory and confirmatory factor analyses using Mplus software (Muthén and Muthén, 2004). Severity threshold parameters for the latent trait of AUD were estimated from a two-parameter logistic model using the BILOG-MG statistical program (Scientific Software International, 2003, see more detail in Saha et al., 2006). Mean values for number of criteria endorsed and validating measures were estimated using SUDAAN (Research Triangle Institute, 2004), a software program that uses Taylor-series linearization to adjust variance estimates for complex design characteristics. SUDAAN multiple linear and logistic regression analyses were used to obtain the net associations of each criterion with the validating measures and to estimate the proportion of variance explained by the simple count of criteria endorsed and an alternate count in which each endorsed criterion was weighted by its severity threshold parameter.

## 3.0 RESULTS

In initial EFA and CFA models (data not shown), factor loadings ranged from 0.699 for legal problems to 0.909 for activities given up, and the model fit statistics all supported a unidimensional model: Comparative Fit Index (CFI) = 0.997, Tucker Lewis Index (TLI) = 0.966, root mean squared error of approximation (RMSEA) = 0.018. As shown in Table 1, the 11 AUD criteria ranged in severity threshold parameter values from 1.49 for drinking more or longer than intended to 3.50 for recurrent legal problems. The severity threshold parameter for each criterion was positively correlated with the total number of positive criteria among individuals endorsing the criterion in question,  $r = .86$  (.98 with legal problems excluded), and negatively correlated with its prevalence,  $r = -.94$ . Accordingly, the mean number of criteria endorsed rose from 3.39 among the 11.5% of past-year drinkers who endorsed drinking more or longer than intended to 7.49 among the 0.9% who reported activities given up in favor of drinking. However, the mean number of positive criteria fell to 5.39 among the 1.0% of past-year drinkers who endorsed legal problems, the criterion with the largest severity threshold parameter. Table 1 also illustrates that individuals endorsing just a few criteria were far more likely to endorse the less severe than the more severe criteria. Among those with just a single positive criterion, there was nearly a 75% likelihood that it was one of the three mildest criteria, and less than a 1% chance that it was among the three most severe criteria. This is to be expected, of course, because the severity parameters are estimated in such a way as to reflect the inverse of their probability of endorsement.

The top panel of Table 2 shows the mean values for the eight outcome or validating measures for individuals who endorsed each of the 11 AUD criteria, listed in order of their severity. For all outcome measures, the values tend to increase directly with criterion severity except for those associated with legal problems, which are lower than expected. For example, average daily volume of ethanol intake increased from a mean of 2.24 ounces among those endorsing drinking more/longer than intended (irrespective of any other criteria they may have endorsed) to 5.86 ounces among those endorsing activities given up, but fell to 3.80 for those endorsing



legal problems. The pattern was reversed for the SF-12 NBMCS score, with lower values of psychological functioning associated with more severe criteria. For all outcomes, the most extreme values were associated with failure to fulfill role obligations and activities given up.

When the net associations of each criterion with the outcomes were estimated in multivariate models that adjusted for age, sex and each other criterion that was endorsed (bottom panel of Table 2), the values of the outcome measures were less positively correlated with the severity ranking of the criteria. Average daily volume, frequency of drinking 5+ drinks, and early initiation of drinking were most strongly associated with endorsement of much time spent drinking, overall frequency of drinking with hazardous use, and largest quantity of drinks with tolerance. Psychological functioning was most strongly associated with use despite physical/psychological problems, and family history density and antisociality were most strongly associated with use despite interpersonal problems. That is, after controlling for other criteria endorsed, associations were often greatest between outcomes and criteria that are inherently linked, such as tolerance and largest quantity of drinks, rather than those with the largest severity threshold parameters.

Figure 1 presents the risk curve for average daily volume of ethanol intake as a function of two alternate measures of severity, the unweighted number of positive criteria (range = 0 to 11) and the severity score that was weighted by the respective severity threshold parameters of the endorsed criteria (range = 0 to 26.07). In order to plot them along a common horizontal axis, the severity scores were multiplied by  $11/26.07$  and rounded to the nearest integer. As is evident in the graphs, the risk curves were closely parallel, with slightly higher values associated with the severity score, especially among individuals endorsing 3–6 criteria or their severity score equivalent. That the curves are very similar is not surprising given the correlation,  $r=.99$ , between the two continuous severity measures. Risk curves for the seven other validating measures were likewise very similar for the two measures, although the shapes of the risk curves varied across outcomes – more linear for the quantity and frequency of drinking measures and more strongly exponential for initiation of drinking prior to age 15 (data not shown).

As shown in Table 3, the slight differences in risk curves between the two dimensional severity measures made no difference in either their correlations with the outcomes or in the proportion of variance that they explained. Either score, or a count of the total number of symptoms endorsed (range=0 to 33) yielded higher values of  $r$  and  $R^2$  than did a dichotomous measure of any AUD. This was also true when the continuous outcome measures (other than NBMCS, which was already normalized) were logged to reduce skew, resulting in multiplicative rather than additive models (data not shown). Exclusion of legal problems, the criterion whose severity parameter appeared most misaligned with the total number of criteria endorsed and with the values of the outcome measures, did not affect the  $r$  and  $R^2$  values.

Table 4 offers information that may help to explain why severity weighting of criteria did not result in a severity scale more strongly associated with outcome measures than the simple count of criteria. Using average daily volume of consumption as an example, Table 4 shows that some of the criteria are operationalized by symptoms that are widely divergent in terms of their apparent severity. For the criterion of tolerance, for example, the average daily volume of ethanol intake among individuals who endorsed the symptom item of drinking the equivalent of a fifth of liquor in a day was twice as great as that among individuals endorsing the symptom item reflecting a decreasing effect of the usual number of drinks, 5.20 versus 2.55 ounces. The variation across symptom items was even more striking in the case of withdrawal; however, it is difficult to interpret the severity associated with any single symptom item because two positive symptoms are required to satisfy the DSM-IV conceptualization of withdrawal as a syndrome. Other criteria marked by a fairly wide range of symptom severity include impaired

control and hazardous use. Since the proportions of individuals endorsing the less severe symptoms of these criteria greatly outweighed those endorsing the more severe symptoms, it can be seen why the overall severity rankings of these criteria were relatively low. At the same time, the data indicate that these criteria have the potential to represent far more severe alcohol problems.

#### 4.0 DISCUSSION

Attempts to validate IRT-based severity parameters for the 11 DSM-IV AUD showed that individuals who endorsed increasingly severe criteria had increasingly large numbers of total criteria endorsed and increasingly high values for a wide variety of AUD correlates (lower values in the case of psychological functioning, where an inverse relationship would be expected). The one exception to this rule was the criterion of recurrent legal problems caused by drinking, which had the highest severity ranking but which was associated with middling levels for total criteria endorsed and the eight outcome/validating measures. This apparent lack of validity with respect to the legal problems criterion is consistent with past research in which deletion of this criterion improved model fit and resulted in more robust IRT parameters, leading to the suggestion that legal problems be dropped as a criterion in future revisions of the DSM (Saha et al., 2006). It may reflect the fact that positive endorsement requires factors that are exogenous to the individual, e.g., apprehension and/or arrest, and whose likelihood may vary according to factors other than AUD severity. Legal problems aside, the severity rankings derived from the IRT analyses were strongly validated by this analysis, closely paralleling variation in AUD correlates that included four aspects of alcohol consumption in addition to mental/psychological functioning, family history of alcoholism, antisociality and early initiation of drinking.

It is noteworthy that the close correlation between severity ranking and outcome values was not maintained after adjusting for other criteria endorsed. This makes an important point, which is that AUD severity is not so much a criterion attribute, *per se*, but rather more an attribute of the individuals who endorse any given criterion. That is, the severity associated with criteria such as failure to fulfill role obligations and important activities given up in favor of drinking reflects the fact that the individuals endorsing those criteria also endorsed many other criteria, i.e., that they had a wide range of alcohol-related problems. In light of this, it is not surprising that an AUD severity scale in which the count of endorsed criteria was weighted by the relative severity of each was no more highly correlated with the outcomes than a scale based on a simple, unweighted count of positive criteria – a finding that represents good news for clinicians, in that the latter is far simpler to calculate in settings where the need for treatment or intervention is evaluated.

Severity weighting would be useful if it could distinguish individuals with just a few positive criteria, all of which were mild, from those with just a few criteria, all of which were severe; however, our data clearly indicated that among individuals with just a few positive criteria, there was a very low likelihood that those criteria would be severe ones. Thus, severity is essentially redundant in light of the extremely high correlation between criterion severity and total number of positive criteria. An interesting way to confirm this interpretation would be to compare severity rankings based on different survey instruments, e.g., those used in the NESARC, NSDUH and other surveys, in terms of their associations with common AUD correlates. If the various rank orderings were equally strongly associated with the validating measures, this would provide even stronger evidence that severity reflects more the number than the nature of the criteria endorsed.

This analysis also illustrated the limitations of a severity scale based solely on a count of criteria endorsed, regardless of whether weighted or unweighted. The severity of some of the criteria

was highly dependent on the symptom items with which they were endorsed. This was particularly the case for the physiological dependence criteria of tolerance and withdrawal, but also true for impaired control and to a lesser degree for hazardous use and time spent drinking. The range of severity indicated by variation in mean volume of intake for individuals endorsing various symptoms parallels IRT-based severity parameters for the symptoms as reported in an earlier analysis (Kahler et al., 2006), and it indicates that some criteria have the potential for representing a far greater level of AUD severity than they do for the majority of individuals endorsing those criteria. It also illustrates that possibly redundant criteria, i.e., those that represent very similar levels of severity, could be better discriminated by choosing operationalizations that led to more dispersion along the latent trait of AUD severity.

Notwithstanding these limitations, it is important to note that all of the continuous severity measures considered in this analysis, i.e., the weighted and simple counts of criteria and a simple count of symptoms, were more highly correlated with the validating measures than was a dichotomous measure of any AUD. This supports earlier findings by Hasin et al. (2006) that a continuous measure of dependence criteria yielded a better model fit than did either a categorical or threshold measure in models predicting family history of alcoholism, early age at onset of drinking and utilization of alcohol treatment. It also adds to arguments supporting inclusion of dimensional as well as categorical diagnostic measures in the upcoming fifth revision of the DSM (DSM-V) (Hasin and Beseler, 2009; Hasin et al., 2006; Helzer et al., 2006, 2007).

Some prior analyses based on IRT models have suggested that a consumption-based criterion based on frequency of drinking 5+/4+ drinks might be a useful addition to represent the lower end of the AUD severity continuum (Saha et al., 2007), although the inclusion of such a measure would substantially affect prevalence estimates and might affect treatment eligibility for some population subgroups (Keyes et al., 2009). Other research has indicated that inclusion of a measure of heavy episodic drinking (HED) did not improve model fit (Beseler et al., 2009) or that it resulted in deviations from linearity in the associations between AUD severity and the outcomes of family history and early initiation of drinking (Hasin and Beseler, 2009). The strong correlations noted between AUD severity scores and frequency of drinking 5+ drinks in this analysis supported the inherent relationship between HED and AUD and suggest that a measure of HED would not be out of place as an indicator of AUD severity. However, the correlations were even stronger for two other consumption measures, average daily volume of ethanol intake and largest quantity of drinks consumed in a single day. In fact, after adjusting for age and sex, the proportion of variance explained by AUD severity was considerably greater for largest quantity of drinks ( $R^2=.39$ ) than for either of the other consumption measures ( $R^2=.30$  for AVD and  $.27$  for frequency of drinking 5+ drinks). This suggests that some dichotomization of largest drinks, e.g., ever consuming 6+, 8+, 10+ or 12+ drinks, might represent another consumption-based criterion with potential for discriminating along the scale of AUD severity. Notably, overall frequency of drinking was less highly correlated with AUD severity than any of the other three consumption measures examined in this analysis. In contrast, Kahler et al. (2009) found that daily drinking represented the most severe of 13 consumption and alcohol problem items in a longitudinal sample spanning adolescence and young adulthood. This suggests that the meaning of different drinking patterns varies across the life span, and perhaps by gender as well, indicating the need to check carefully for differential item functioning (DIF) before including consumption measures, even those with built-in gender adjustments such as frequency of drinking 5+/4 drinks, as potential AUD criteria.

The findings of this study must be interpreted with some caution in light of the fact that the measures used to validate AUD severity were, like the symptom items used to measure severity, based on self report. If the likelihood of completely and accurately reporting AUD symptom



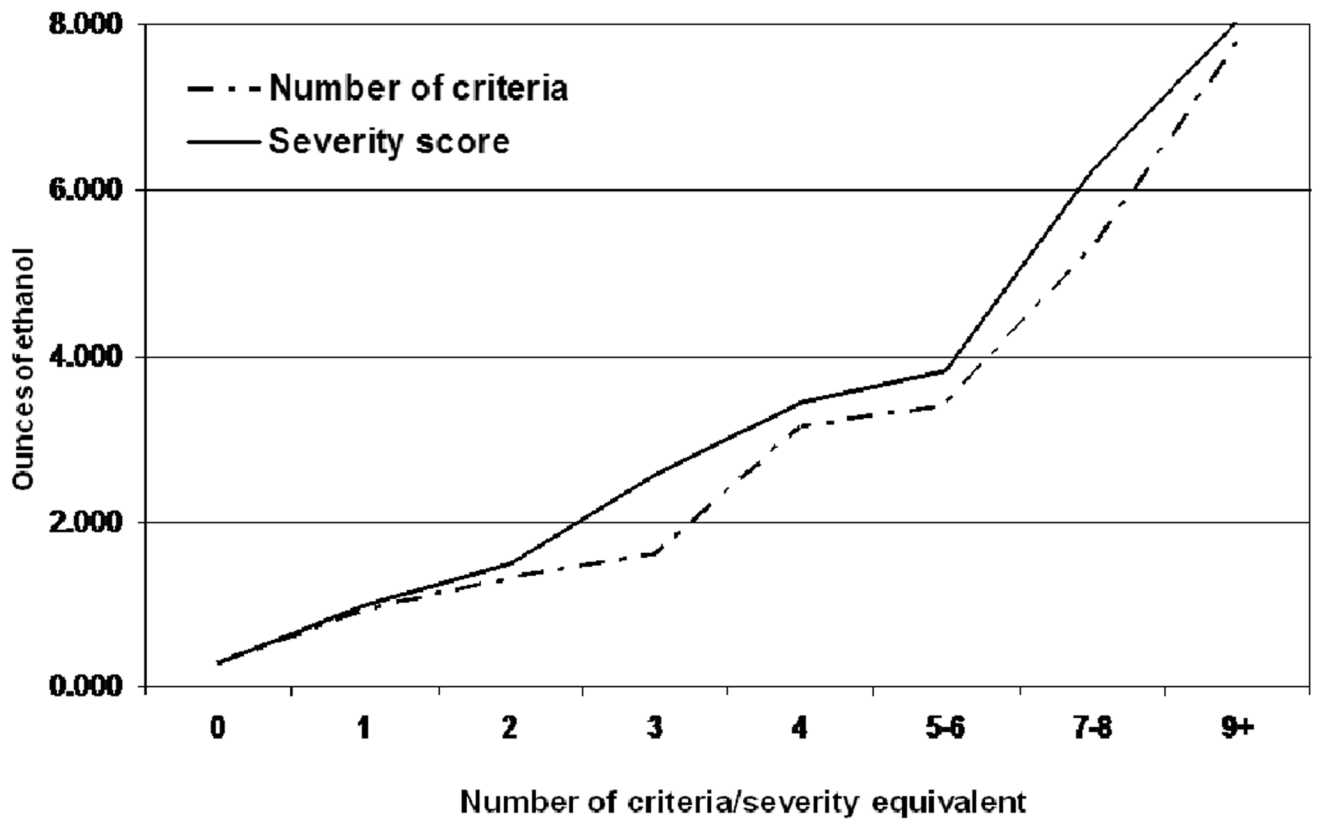
item indicators was correlated with the likelihood of completely and accurately reporting consumption, psychological functioning, family history of alcoholism, symptoms of antisociality and age at first drink, this could lead to overestimation of the correlation between severity and the outcome measures. External measures, e.g., collateral reports or records-based information, would provide a more unbiased assessment of validity. Likewise, the prognostic value of the severity scales might better be tested prospectively, as could be done using data from Wave 2 of the NESARC. However, there is no reason to think that externally or prospectively derived outcome measures would lead to different conclusions regarding the validity of the relative severity ranking of the AUD criteria. Similarly, whereas the slopes for the risk curves linking AUD severity and the various outcomes might be slightly exaggerated as a result of using self-reported outcomes measures, any bias of this sort should not affect the comparison of the weighted and unweighted AUD severity risk curves. Thus, we would argue that the results of this paper are sufficiently robust to confirm the importance of including a dimensional measure of AUD severity in the DSM-V and the utility of a simple symptom or criterion count for this purpose.

## REFERENCES

- American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders. 4th Edition. Washington, D.C: American Psychiatric Association; 1994.
- Beseler, CL.; Shmulewitz, E.; Aharonovich, E.; Hasin, DS. Alcohol. Clin. Exper. Res. Vol. 33. 2009. DSM-IV alcohol abuse and dependence: An IRT analysis in Israeli household residents; p. 65A
- Bohannon RW, Maljanian R, Landes M. Test-retest reliability of short form (SF)-12 component scores of patients with stroke. *Int. J. Rehabil. Res* 2004;27:149–150. [PubMed: 15167113]
- Butler SF, Redondo JP, Fernandez KC, Villapiano A. Validation of the Spanish Addiction Severity Index Multimedia Version (S-ASI-MV). *Drug Alcohol Depend* 2009;99:18–27. [PubMed: 18718727]
- Daepfen J-B, Burnand B, Schnyder C, Bonjour M, Pecoud A, Yersin B. Validation of the Addiction Severity Index in French-speaking alcoholic patients. *J. Stud. Alcohol* 1995;57:585–590. [PubMed: 8913988]
- Donovan D, Kivlahan DR, Doyle SR, Longabaugh R, Greenfield S. Concurrent validity of the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT zones in defining levels of severity in the COMBINE study. *Addiction* 2006;101:1696–1704. [PubMed: 17156168]
- Edwards G, Gross MM. Alcohol dependence: provisional description of a clinical syndrome. *Br. Med. J* 1976;1:1058–1061. [PubMed: 773501]
- Grant, BF.; Dawson, DA.; Hasin, DS. The Alcohol Use Disorders and Associated Disabilities Interview Schedule - Version for DSM-IV (AUDADIS-IV). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2001.
- Grant BF, Dawson DA, Stinson FS, Chou PS, Pickering RP. The Alcohol Use Disorder and Associated Disabilities Schedule (AUDADIS): Reliability of alcohol and drug modules in a general population sample. *Drug Alcohol Depend* 2003b;39:7–16.
- Grant BF, Harford TC, Dawson DA, Chou PS, Pickering RP. The alcohol use disorder and associated disabilities schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. *Drug Alcohol Depend* 1995;39:37–44. [PubMed: 7587973]
- Grant, BF.; Moore, TC.; Shepard, J.; Moore, T. Source and Accuracy Statement: Wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Bethesda MD: National Institute on Alcohol Abuse and Alcoholism; 2003a.
- Grant, BF.; Peterson, A.; Dawson, DA.; Chou, SP. Source and Accuracy Statement for the National Longitudinal Alcohol Epidemiologic Survey: Wave 1. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism; 1994.
- Harford TC, Yi H-Y, Faden V, Chen CM. The dimensionality of DSM-IV alcohol use disorders among adolescent and adult drinkers and symptom patterns by age, gender and race-ethnicity. *Alcohol. Clin. Exp. Res* 2009;33:868–878. [PubMed: 19320629]
- Hasin DS, Beseler CL. Dimensionality of lifetime alcohol abuse, dependence and binge drinking. *Drug Alcohol Depend* 2009;101:53–61. [PubMed: 19095379]

- Hasin DS, Liu H, Alderson D, Grant BF. DSM-IV alcohol dependence: a categorical or dimensional phenotype? *Psychol. Med* 2006;36:1695–1705. [PubMed: 17038207]
- Helzer JA, Bucholz KK, Gossop M. A dimensional option for the diagnosis of substance dependence in DSM-V. *Int. J. Methods Psychiatr. Res* 2007;16:S24–S33.
- Helzer JE, van denBrink W, Guth SE. Should there be both categorical and dimensional criteria for the substance use disorders in DSM-V? *Addiction* 2006;101:17–22. [PubMed: 16930157]
- Kahler CW, Hoepfner BB, Jackson KM. A Rasch model analysis of alcohol consumption and problems across adolescence and young adulthood. *Alcohol. Clin. Exp. Res* 2009;33:1–11. [PubMed: 18828798]
- Kahler CW, Strong DR. A Rasch model analysis of DSM-IV alcohol abuse and dependence items in the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcohol. Clin. Exp. Res* 2006;30:1165–1175. [PubMed: 16792564]
- Kahler CW, Strong DR, Stuart GL, Moore TM, Ramsey SE. Item functioning of the alcohol dependence scale in a high risk sample. *Drug Alcohol Depend* 2003;72:183–192.
- Keyes KM, Geier T, Brant BF, Hasin DS. Influence of a drinking quantity and frequency measure on the prevalence and demographic correlates of DSM-IV alcohol dependence. *Alcohol. Clin. Exp. Res* 2009;33:761–771. [PubMed: 19298332]
- King JT Jr, Horowitz MB, Kassam AB, Jonas H, Roberts MS. The short form-12 and the measurement of health status in patients with cerebral aneurysms: performance, validity, and reliability. *J. Neurosurg* 2005;102:489–494. [PubMed: 15796384]
- Lima AF, Pechansky F, Fleck MP, de Boni R. Association between psychiatric symptoms and severity of alcohol dependence in a sample of Brazilian men. *J. Nerv. Ment. Dis* 2005;193:126–130. [PubMed: 15684915]
- McLellan AT, Luborsky L, Cacciola J, Griffith JE, Evans F, Barr HL, O'Brien CP. New data from the Addiction Severity Index: reliability and validity in three centers. *J. Nerv. Ment. Dis* 1985;173:412–423. [PubMed: 4009158]
- Muthén, BO.; Muthén, LK. *Mplus: Statistical Analysis with Latent Variables (Version 3.01)*. Los Angeles, CA: Muthén & Muthén, Inc; 2004.
- Proudfoot H, Baillie AJ, Teeson M. The structure of alcohol dependence in the community. *Drug Alcohol Depend* 2006;81:21–26. [PubMed: 16005578]
- Research Triangle Institute. *Software for survey data analysis (SUDAAN)*. Research Triangle Park, NC: Version 9.0. Research Triangle Institute; 2004.
- Resnick B, Nahm ES. Reliability and validity testing of the revised 12-item Short-Form Health Survey in older adults. *J. Nurs. Meas* 2001;9:151–161. [PubMed: 11696939]
- Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol. Med* 2006;36:931–941. [PubMed: 16563205]
- Saha TD, Stinson FS, Grant BF. The role of alcohol consumption in future classifications of alcohol use disorders. *Drug Alcohol Depend* 2007;89:82–92. [PubMed: 17240085]
- Salyers MP, Bosworth HB, Swanson JW, Lamb-Pagone J, Osher FC. Reliability and validity of the SF-12 health survey among people with severe mental illness. *Med. Care* 2000;38:1141–1150. [PubMed: 11078054]
- Saunders JB, Aasland OG, Babor TF, de la, Fuente JR, Grant M. Development of the Alcohol Use Disorder Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption II. *Addiction* 1993;88:791–804. [PubMed: 8329970]
- Skinner HA, Horn JL. Alcohol dependence syndrome: measurement and validation. *J. Abnorm. Psychol* 1984;91:199–209. [PubMed: 7096790]
- Du Toit, M., editor. *Scientific Software Internationala. (SSI). Item Response Theory (IRT) from SSI*. Lincolnwood, IL: Scientific Software International; 2003.
- Stockwell T, Hodgson R, Edwards G, Taylor C, Rankin H. The development of a questionnaire to measure severity of alcohol dependence. *Br. J. Addiction* 1979;74:79–87.
- Stoltenberg SF, Mudd SA, Blow FC, Hill EM. Evaluating measures of family history of alcoholism: density versus dichotomy. *Addiction* 1998;93:1511–1520. [PubMed: 9926555]

- Thomas BA, McCambridge J. Comparative psychometric study of a range of hazardous drinking measures administered online in a youth population. *Drug Alcohol Depend* 2008;96:121–127. [PubMed: 18406079]
- Ware, JE.; Kosinski, M.; Turner-Bowker, DM.; Gandek, B. *How to Score Version 2 of the SF-12® Health Survey (With a Supplement Documenting Version 1)*. Lincoln, RI: QualityMetric Incorporated; 2002.
- Wood LD, Sobell LC, Sobell MB, Dornheim L, Agrawal S. Nondaily drinkers score higher on the Alcohol Dependence Scale than daily drinkers. *Addict. Behav* 2003;28:347–352. [PubMed: 12573684]
- Zimmerman M, Mattia JI. The psychiatric diagnostic screening questionnaire development, reliability and validity. *Compr. Psychiatry* 2001;42:175–189. [PubMed: 11349235]



**Figure 1.**  
Average daily ethanol intake as a function of number of criteria endorsed and severity scale equivalent: U.S. past-year drinkers 18 years of age and over

**Table 1**

Severity threshold parameters for specific DSM-IV AUD criteria, total numbers of AUD criteria endorsed by individuals who endorsed each criterion, and probability of criterion endorsement among drinkers endorsing a total of one, two or three criteria: U.S. past-year drinkers 18 years of age and older

Criterion	Severity Threshold parameter	Mean number of AUD criteria endorsed*	% endorsing criterion	Probability of criterion endorsement among those endorsing:		
				One criterion	Two criteria	Three criteria
Drank more/longer than intended	1.49 (0.02)	3.39 (0.06)	11.5 (0.3)	.226 (.010)	.483 (.017)	.734 (.021)
Impaired control (stop/cut down)	1.80 (0.02)	3.25 (0.07)	9.8 (0.3)	.255 (.010)	.388 (.016)	.479 (.022)
Hazardous use	1.81 (0.02)	3.28 (0.07)	9.6 (0.3)	.251 (.011)	.374 (.017)	.462 (.023)
Withdrawal	1.91 (0.02)	3.99 (0.08)	7.1 (0.2)	.093 (.006)	.272 (.014)	.432 (.021)
Tolerance	1.92 (0.02)	3.79 (0.07)	7.5 (0.2)	.131 (.008)	.257 (.013)	.441 (.025)
Use despite physical/psychological problems	2.21 (0.03)	5.05 (0.10)	4.3 (0.2)	.022 (.003)	.114 (.012)	.189 (.018)
Time spent drinking	2.50 (0.03)	5.90 (0.10)	2.7 (0.1)	.004 (.001)	.036 (.006)	.089 (.011)
Use despite interpersonal problems	2.60 (0.03)	5.74 (0.14)	2.6 (0.1)	.012 (.001)	.042 (.006)	.099 (.014)
Failure to fulfill role obligations	3.08 (0.04)	7.33 (0.18)	1.2 (0.1)	.000 (.000)	.007 (.005)	.013 (.004)
Activities given up	3.25 (0.05)	7.49 (0.17)	0.9 (0.1)	.001 (.001)	.004 (.002)	.016 (.005)
Legal problems	3.50 (0.06)	5.39 (0.23)	1.0 (0.1)	.006 (.001)	.022 (.005)	.045 (.011)

Note: Figures in parentheses are standard errors of estimates.

\* Correlation with severity threshold parameter = 0.86 including legal problems and 0.98 excluding legal problems



Table 2

Mean values of selected correlates of AUD among U.S. past-year drinkers 18 years of age and older who endorsed specified DSM-IV AUD criteria

	Average daily volume (oz. ethanol)		Overall freq. of drinking (days/year)		Largest quantity drinks		Freq. drinking 5+ drinks (days/year)		SF-12 NBMCS Range: 0-100		Family history Density Range: 0-2		No. ASPD Symptoms Range: 0-33		% with age at 1st drink <15	
	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)
<i>Unstandardized estimates</i>																
Drank more/longer than intended	2.24	(0.09)	151.2	(2.8)	10.4	(0.2)	81.9	(2.6)	49.8	(0.2)	0.32	(0.01)	5.7	(0.2)	13.7	(0.8)
Impaired control (stop/cut down)	2.24	(0.10)	155.8	(3.6)	9.4	(0.2)	76.5	(2.9)	49.7	(0.3)	0.32	(0.01)	5.4	(0.2)	14.4	(0.9)
Hazardous use	2.30	(0.08)	160.4	(3.0)	10.7	(0.2)	85.8	(2.8)	51.0	(0.2)	0.31	(0.01)	5.7	(0.1)	13.5	(0.9)
Withdrawal	2.50	(0.12)	149.1	(3.4)	10.7	(0.3)	87.2	(3.4)	48.2	(0.3)	0.36	(0.01)	6.5	(0.2)	17.3	(1.2)
Tolerance	2.92	(0.13)	156.9	(3.4)	12.3	(0.3)	97.8	(3.3)	49.4	(0.3)	0.33	(0.01)	6.1	(0.2)	15.1	(1.0)
Use despite physical/ psychological problems	3.60	(0.18)	183.6	(5.1)	12.6	(0.3)	121.6	(5.1)	46.0	(0.5)	0.43	(0.01)	7.8	(0.3)	21.5	(1.7)
Time spent drinking	4.80	(0.28)	205.0	(5.4)	15.0	(0.5)	149.3	(6.1)	46.4	(0.5)	0.41	(0.02)	8.6	(0.3)	26.8	(2.1)
Use despite interpersonal problems	4.42	(0.26)	188.5	(5.8)	14.6	(0.5)	137.6	(6.2)	46.4	(0.6)	0.47	(0.02)	9.6	(0.4)	25.3	(2.2)
Failure to fulfill role obligations	5.22	(0.41)	210.2	(9.2)	15.1	(0.7)	166.0	(10.2)	42.8	(0.9)	0.53	(0.04)	10.9	(0.6)	31.0	(3.6)
Activities given up	5.86	(0.47)	225.8	(10.2)	15.4	(0.8)	180.6	(10.6)	43.5	(1.0)	0.50	(0.03)	10.8	(0.8)	32.3	(3.6)
Legal problems	3.80	(0.42)	168.0	(8.1)	13.3	(0.9)	123.4	(8.6)	49.1	(1.0)	0.41	(0.04)	8.6	(0.7)	25.3	(3.5)
<i>Adjusted* multivariate models results</i>																
Drank more/longer than intended	0.305	.000	34.9	.000	2.02	.000	19.5	.000	-0.610	.032	0.020	.052	0.709	.000	0.004	.969
Impaired control	0.471	.000	40.1	.000	1.10	.000	15.8	.000	-1.150	.000	0.043	.000	0.501	.002	0.252	.016
Hazardous use	0.494	.000	44.1	.000	2.76	.000	27.5	.000	0.616	.025	0.035	.002	0.858	.000	0.038	.730
Withdrawal	0.152	.136	8.9	.014	0.93	.000	6.8	.030	-1.856	.000	0.047	.001	0.988	.000	0.362	.002
Tolerance	0.980	.000	28.1	.000	3.79	.000	29.9	.000	-0.691	.044	0.010	.416	0.571	.003	-0.004	.975
Use despite physical/ psychological problems	0.617	.000	23.2	.000	1.05	.002	24.8	.000	-3.513	.000	0.090	.000	1.044	.000	0.312	.034
Time spent drinking	1.682	.000	28.6	.000	2.14	.000	36.7	.000	-0.837	.195	0.001	.981	0.815	.031	0.577	.000
Use despite interpersonal problems	1.114	.000	8.5	.191	2.30	.000	22.6	.001	-1.421	.022	0.099	.000	2.636	.000	0.359	.033
Failure to fulfill role obligations	0.443	.296	-5.6	.651	-0.22	.769	12.6	.246	-3.036	.003	0.071	.075	1.721	.016	0.228	.292
Activities given up	1.187	.016	13.8	.237	-0.16	.851	30.1	.010	-1.404	.167	0.025	.518	1.181	.114	0.223	.282
Legal problems	0.582	.161	-5.2	.501	0.98	.293	12.5	.130	0.624	.489	0.032	.363	1.152	.056	0.380	.093

Note: Figures in parentheses are standard errors of estimates. Shaded cells denote highest values (lowest for NBMCS).

\* Adjusted for age, sex and other criteria endorsed

Associations of selected measures of overall AUD severity with selected correlates of AUD among U.S. past-year drinkers 18 years of age and older

**Table 3**

	Average daily volume	Overall freq. of drinking	Largest quantity drinks	Freq. drinking 5+ drinks	SF-12 NBMCS	Family history density	No. ASPD symptoms	Age 1 <sup>st</sup> drink <15
<b>Bivariate correlation coefficients (r)</b>								
No. criteria	.52	.30	.55	.49	-.16	.18	.36	.15
No. criteria, excl. legal problems	.52	.30	.55	.49	-.16	.18	.36	.15
Severity score	.53	.29	.54	.49	-.17	.18	.36	.15
Severity score, excl. legal problems	.53	.29	.54	.49	-.17	.18	.36	.15
No. symptoms	.54	.28	.55	.50	-.17	.18	.35	.16
Any AUD (dichotomous)	.39	.25	.49	.40	-.10	.13	.29	.11
<b>R square from multivariate regression models adjusted for age and sex</b>								
No. criteria	.30	.16	.39	.27	.05	.06	.16	.02
No. criteria, excl. legal problems	.30	.16	.39	.27	.05	.06	.16	.02
Severity score	.30	.15	.38	.27	.05	.06	.16	.02
Severity score, excl. legal problems	.30	.15	.38	.27	.05	.06	.16	.02
No. symptoms	.30	.15	.38	.27	.05	.06	.16	.02
Any AUD (dichotomous)	.16	.12	.31	.18	.03	.05	.12	.01

**Table 4**

Average daily volume of ethanol intake among individuals who endorsed selected AUD symptom items: U.S. past-year drinkers 18 years of age and older

Criterion/symptom item	Average daily volume	% endorsing symptom
<b>Drank more/longer than intended</b>		
Drank more than meant to	2.32 (0.10)	9.6 (0.3)
Drank longer than meant to	2.50 (0.10)	7.8 (0.3)
<b>Impaired control (stop/cut down)</b>		
More than once wanted to stop/cut down	2.24 (0.10)	9.5 (0.3)
More than once tried and failed to stop/cut down	3.86 (0.30)	1.9 (0.1)
<b>Hazardous use</b>		
Drove car while drinking	2.26 (0.09)	6.9 (0.3)
Drove vehicle after having too much to drink	2.76 (0.13)	4.4 (0.2)
Got into risky situations during/after drinking	3.39 (0.22)	2.8 (0.2)
<b>Withdrawal</b>		
Sick to stomach	1.87 (0.09)	9.0 (0.3)
Trouble sleeping	2.12 (0.15)	4.3 (0.2)
Restless	2.66 (0.18)	3.4 (0.2)
Sweating/rapid heartbeat	2.96 (0.21)	3.1 (0.1)
Nervous/anxious	3.91 (0.28)	2.0 (0.1)
Shaking	4.22 (0.32)	1.7 (0.1)
Hallucinations	5.14 (0.59)	0.6 (<0.1)
Seizures	5.51 (1.89)	0.1 (<0.1)
<b>Relief of withdrawal</b>		
Took drink/medicine to get over aftereffects	3.51 (0.22)	2.4 (0.1)
Took drink/medicine to prevent aftereffects	4.11 (0.22)	1.4 (0.1)
<b>Tolerance</b>		
Usual no. drinks had less effect than once did	2.55 (0.15)	5.0 (0.4)
Had to drink more to get effect desired	3.50 (0.24)	3.1 (0.2)
Increased drinking because former amount didn't give same effect	3.93 (0.28)	2.1 (0.1)
Drank equivalent of a fifth in a day	5.20 (0.29)	2.8 (0.1)
<b>Continued use despite physical/psychological problems</b>		
Use despite health problems	4.14 (0.26)	2.3 (0.1)
Use despite prior blackout	4.42 (0.30)	2.2 (0.1)
Use despite psychological problems	4.55 (0.32)	1.5 (0.1)
<b>Time spent drinking</b>		
Time spent getting over effects of drinking	4.33 (0.34)	1.0 (0.1)
Time spent drinking	5.27 (0.32)	2.3 (0.1)
<b>Continued use despite interpersonal problems</b>		
Got into physical fights while/after drinking	4.70 (0.35)	1.6 (0.1)

Criterion/symptom item	Average daily volume	% endorsing symptom
Caused problems with family or friends	5.36 (0.39)	1.4 (0.1)
<b>Failure to fulfill role obligations</b>		
Drinking/aftereffects interfered with care of home/family	5.32 (0.48)	0.9 (0.1)
Job/school problems as a result of drinking/aftereffects	5.64 (0.67)	0.6 (0.1)
<b>Activities given up</b>		
Gave up/cut down activities that interested/gave pleasure to you	6.14 (0.59)	0.6 (0.1)
Gave up/cut down activities important to you	6.34 (0.53)	0.8 (0.1)
<b>Legal problems</b>		
Arrested/held at police station/other legal problems	3.80 (0.43)	1.0 (0.1)

Note: Figures in parentheses are standard errors of estimates