Alcohol and Alcoholism, 2015, 50(4) 413–419
doi: 10.1093/alcalc/agv018
Advance Access Publication Date: 13 March 2015
Article

OXFORD

Article

Assessing the Validity of Participant-Derived Compared to Staff-Derived Values to Compute a Binge Score

Sarah L. Lake[1], Nathalie Hill-Kapturczak[1], Yuanyuan Liang[2], John D. Roache[1], Jillian Mullen[1], Tara E. Karns[1], and Donald M. Dougherty[1,*]

[1]Psychiatry Department, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA, and [2]Department of Epidemiology and Biostatistics, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

*Corresponding author: Division of Neurobehavioral Research, Psychiatry Department, The University of Texas Health Science Center at San Antonio, NRLC MC 7793, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA. Tel.: +1-210-5672745; Fax: +1-210-5672748; E-mail: doughertyd@uthscsa.edu

Received 1 October 2014; Revised 18 February 2015; Accepted 19 February 2015

Abstract

Aims: This study examined the validity of two methods of classifying binge drinkers.
Methods: Adult drinkers ($n = 166$) completed the Alcohol Use Questionnaire (AUQ) and a Timeline Followback (TLFB) interview to characterize drinking during the past 28 days. Using Townshend and Duka's (2005) recommendations, answers on three AUQ items (average drinks per hour, number of times drunk within the prior 6 months and percentage of times drunk when drinking) were used to derive a binge score that was then used to classify drinkers as Binge, Non-Binge and Unclassifiable. Two methods for calculating binge scores were compared: (a) Participant-derived, using participants' answers on the 3 AUQ items; and (b) Staff-derived, staff used TLFB interview information to answer the 3 AUQ items. Additionally, Participant- and Staff-derived classifications were used to predict future drinking behaviors assessed by a second TLFB interview.
Results: Participant- and Staff-derived binge scores had a low concordance rate. Staff-derived classifications were better than Participant-derived classifications at predicting future binge drinking behavior and identifying group differences in drinking behavior reported during the second TLFB interview (average drinks per hour, number of times drunk within the prior 6 months, and percentage of times drunk when drinking).
Conclusions: Classifying drinkers using staff-guided TLFB interview methods instead of self-reported participant generalizations of typical drinking habits better relates to real-world drinking. Classification schemes that rely on dichotomous categorization of drinkers (Binge vs. Non-Binge) may be missing individuals who engage in harmful patterns of drinking. A continuous scale or index characterizing problematic drinking may be more useful.

INTRODUCTION

Binge drinking, the most common pattern of excessive alcohol use (CDC, 2012), is associated with physical and psychological health problems such as liver cirrhosis, cancers, sexually transmitted diseases, stroke and social problems such as interpersonal violence and drunk driving (Shultz *et al.*, 1991; NIAAA, 2000; Carlson *et al.*, 2010; Orchowski *et al.*, 2012; CDC, 2014). The risk of accidents, unintentional harm, suicide, aggression and their associated injuries are greater

among binge drinkers than chronic heavy drinkers (Borges and Hansen, 1993; Duncan, 1997; Oei and Morawska, 2004). Binge drinking was responsible for about half of the 87,798 alcohol-attributable deaths (per year) in the United States from 2006 through 2010 (Stahre *et al.*, 2014). Given these consequences, it is critically important to accurately measure binge drinking and identify binge drinkers for both clinical research and treatment purposes.

The definition of binge drinking, however, varies depending on the study (Courtney and Polich, 2009). Studies using the number of standard drinks per drinking occasion to define binge drinking have differed in the number of drinks necessary to meet criteria, [e.g. ≥4 for women or ≥6 for men (e.g. Olthuis *et al.*, 2011), ≥5 for both sexes (e.g. Naimi *et al.*, 2003) and ≥6 for both sexes (e.g. Charles *et al.*, 2011; Bauer and Ceballos, 2014)]. Such definitions omit rates of drinking, and the frequency with which these episodes occur (Gmel *et al.*, 2011). The National Institute of Alcohol Abuse and Alcoholism defines binge drinking episodes for the typical adult as consumption ≥4 drinks for women or ≥5 drinks for men within 2 h (NIAAA, 2004). While this definition defines a specific binge episode, it does not classify an individual as being a 'binge drinker.' Presumably, such a classification has to take into account the frequency of binge episodes. Binge drinking patterns are distinct from both moderate drinking and chronic heavy alcohol drinking. Moderate drinkers consume much less alcohol per episode than binge drinkers (Courtney and Polich, 2009), and chronic heavy alcohol users consume large quantities per episode but do not have significant periods of abstinence between drinking episodes (Auerbach and Collins, 2006; Jackson *et al.*, 2006). Furthermore, *intermittent patterns* of heavy alcohol consumption over fewer drinking days, rather than the number of drinking days, may underlie the heightened mortality of binge drinkers (Oei and Morawska, 2004).

Definitions of a 'binge drinker' based on the frequency of binges have also varied greatly. For example, the Substance Abuse and Mental Health Services Administration has defined individuals with at least one binge episode (using the NIAAA definition) in the past 30 days as binge drinkers (SAMHSA, 2007, 2012). Other researchers have used definitions such as one binge in the past 2 weeks (Wechsler *et al.*, 1995), one per week for the previous 3 months (e.g. Beseler *et al.*, 2012; Aston *et al.*, 2014), or one in the past year (Cranford *et al.*, 2006). In summary, three factors appear to be commonly considered when defining what constitutes a binge drinker: quantity of alcohol consumed, rate of drinking and frequency of binge episodes.

Townshend and Duka (2005) used the last three questions from the Alcohol Use Questionnaire (AUQ; Mehrabian and Russell, 1978,), which incorporates these three factors, to develop an equation and a resulting binge score to classify different types of drinkers. These three AUQ questions ask respondents to provide a generalized estimate of typical drinking patterns, including: (a) average drinks per hour; (b) number of times drunk within the prior 6 months and (c) percentage of times drunk when drinking. Based on the binge scores Townshend and Duka (2005) classified the top third of their sample as binge drinkers and the bottom third as non-binge drinkers, with the minimum and maximum for each group used as cut-off scores. Those who fell between the cut offs were considered unclassifiable. Binge scores derived from the AUQ have been reported to predict later problems with alcohol (Townshend and Duka, 2005).

In the current study, we compared two ways of calculating this binge score to examine its validity. Specifically, we asked subjects to self-report their responses to the AUQ and we asked trained research staff to answer the same three questions, based on information they collected during calendar-based Timeline Followback (TLFB; Sobell and Sobell, 1992) interviews. We hypothesized that: (a) binge scores derived from self-reported generalizations of drinking on the AUQ would not be convergent with binge scores derived from TLFB interviews; and (b) AUQ classifications of types of drinkers (Non-Binge, Binge or Unclassified drinkers) using binge scores based on the TLFB interviews would better predict future drinking than classifications based upon self-reported general drinking.

METHODS

Participants

A total of 166 healthy adult drinkers (102 men, 64 women) participated in the present study. Participants were recruited from the San Antonio, Texas area by radio, internet, television and print advertisements. Respondents completed an initial telephone interview that included demographic characteristics, substance use, medical health and psychiatric history. Individuals ($n = 1465$) who met initial eligibility criteria: aged 26–54 years, reported typically drinking 1–4 times per week, had no current or chronic drug use and no clinically significant medical or psychiatric conditions were invited to the laboratory for more extensive screening. This screening included a more detailed substance use history, psychiatric screening for clinical disorders using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I/NP; First *et al.*, 2001), intelligence screening using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), urinalysis for recent drug use (cocaine, THC, opiates, barbiturates, benzodiazepines and methamphetamine) or pregnancy and a physical examination by a physician's assistant. Exclusion criteria included body mass index >35, IQ < 70, Axis I disorder (including alcohol or other substance dependence), positive drug or pregnancy test at the time of screening or a medical condition that would contraindicate alcohol use (e.g. diabetes or liver disease). Participants received $70 for the in-person laboratory screening day and $35 per week for 4 weeks to complete weekly TLFB interviews. Of those screened, 370 individuals were determined eligible and 166 of these elected to participate in the study. All participants completed written informed consent for the study as approved by the Institutional Review Board at The University of Texas Health Science Center at San Antonio.

Methods of binge score computation

Binge scores were calculated using Townshend and Duka's (2005) method, which uses the last three items from the AUQ (Mehrabian and Russell, 1978) to characterize patterns of drinking: items 10 (average drinks per hour), 11 (how many times drunk in the previous 6 months) and 12 (percentage of times getting drunk when drinking). Answers are inserted into an equation to yield a *binge score*: **(4 × (Item 10)) + Item 11 + (0.2 × (Item 12))**. Scores are then used to classify drinkers as Non-Binge (score ≤ 16), Binge (score ≥ 24) or Unclassified (score >16 but < 24). We calculated binge scores from these three items in two ways: (a) using values self-reported by participants on the AUQ; and (b) using information gathered from TLFB (Sobell and Sobell, 1992) interviews, which was used by the research staff to answer the three AUQ questions. Each method is described below.

Participant-derived binge scores

At study entry, participants completed the AUQ (Mehrabian and Russell, 1978) which consists of 12 questions that assess habitual alcohol consumption over the previous 6 months. The last three questions (average drinks per hour, how many times drunk in the previous 6 months and percentage of times getting drunk when drinking) of the 12 questions were used to calculate the binge scores using Townshend

and Duka's (2005) equation. Derived values from this method are referred to as Participant-derived binge scores.

Staff-derived binge scores

Alternatively, study staff used information gathered from TLFB interviews (Sobell and Sobell, 1992) to answer these same three AUQ questions and calculate a binge score based on Townshend and Duka's (2005) equation. The TLFB is a calendar-based semi-structured interview where trained research assistants prompt participants to recall their alcohol consumption by date, using memory aids, such as holidays, to help them with recall. Participants are asked the quantity of alcohol consumed, specific brand of alcohol and hours over which a drinking event occurred for the 28 days before study entry. Drinks are then converted into standard units based on alcohol-by-volume percentages. The TLFB is psychometrically valid in a range of settings for assessing daily self-reported alcohol consumption (Sobell and Sobell, 1992; Sobell *et al.*, 1996), and has adequate test-retest reliability (Sobell *et al.*, 1988).

Study staff used the TLFB completed at study entry (TLFB-1) to calculate values for the same 3 AUQ items used in the equation above to derive binge scores: (a) average drinks per hour [computed by summing all drinks in the prior month and dividing by the total number of drinking hours]; (b) number of times drunk in the past six months [computed by summing the number of binges in the prior month (using NIAAA's definition of ≥4 or ≥5 units within 2 h for women and men, respectively as this would typically result in a blood alcohol level of 0.08 g/dl, the legal definition of intoxication in the United States) and multiplying it by 6] and (c) percentage of times getting drunk when drinking [computed by summing the total number of binges in the prior month and dividing by total number of drinking episodes]. Thus, one month of TLFB was extrapolated to estimate each participant's typical drinking pattern. Values from these items are referred to as Staff-derived binge scores.

Assessment of drinking after study entry

In addition to TLFB-1 (completed at study entry), additional TLFB interviews were completed weekly for 4 weeks during participation in a larger study (TLFB-2; Dougherty *et al.*, 2015). TLFB-2 data were gathered to determine how well Participant- and Staff-derived binge scores related to future drinking behavior.

Data analysis

Demographic characteristics of the two sexes were compared using *t*-tests or Chi-squared tests for continuous or categorical variables, respectively. Two steps of analyses were performed. In the first step of analyses, convergent validity between Participant- and Staff-derived values on the three AUQ items (i.e. average drinks per hour, number of times drunk and percentage drunk when drinking) was examined using Pearson's correlation coefficients and corresponding 95% confidence intervals (CI). Agreement between Participant-derived and Staff-derived binge scores was assessed by testing the null hypothesis from a simple linear regression such that the intercept is 0 and the slope is 1, and agreement between Participant-derived and Staff-derived drinker categorizations (i.e. Binge, Non-Binge or Unclassified drinkers) was determined by a weighted κ statistic. Simple linear regression models were also used to examine the variance accounted for in average drinks per hour, number of times drunk and percentage drunk when drinking by Participant-derived binge scores vs. Staff-derived binge scores.

In the second step, a multivariate analysis of variance (MANOVA) with three independent variables (Participant-derived drinking classifications, Staff-derived drinking classifications and their interaction) was used to examine whether Staff-derived drinking classifications compared to Participant-derived drinking classifications at study entry would better predict future drinking at TLFB-2 (average drinks per hour, number of times drunk and percentage drunk when drinking). Post-hoc Tukey's tests were used to further examine significant main effects. TLFB-2 values were square root-transformed to maintain normality assumptions (Tabachnik and Fidell, 2007). Lastly, Pearson's correlation coefficients were calculated between each of the three AUQ items derived from TLFB-1 and TLFB-2.

Finally, Receiver Operator Characteristic (ROC) curves were calculated to compare Participant- and Staff-derived binge scores in predicting binge behaviors from TLFB-2. Specifically, participants were rank ordered and split into quartiles, using the top (frequent bingers) and bottom (infrequent bingers) quartile as the measure of binge behavior from TLFB-2. The area under the curve (AUC) was used as a measure of predictive power of the binge scores.

RESULTS

Participant characteristics

Characteristics of the 166 participants appear in Table 1. Compared to men, women reported having slightly more binges in the month before study entry and being drunk more often when drinking.

Convergent validity between participant- and staff-derived binge scores

Using Pearson's correlation coefficients, we examined the correlation between the Participant and Staff-derived responses to the three items from the AUQ used to calculate binge scores based on Townshend and Duka's (2005) recommendations: drinks per hour, number of times drunk and percentage drunk when drinking (results not displayed). Significant but modest correlations were observed for the average number of drinks per hour ($r = 0.341$, 95% CI = [0.199, 0.469], $P < 0.001$) and percentage of times getting drunk when drinking ($r = 0.224$, 95% CI = [0.074, 0.364], $P < 0.004$). Participant-derived reports of the number of times drunk in the previous six months

Table 1. Demographic characteristics at study entry

	Women ($n = 64$)		Men ($n = 102$)		Total ($n = 166$)		*P*
	M	*SD*	*M*	*SD*	*M*	*SD*	
Age	31.25	8.25	30.96	8.77	31.07	8.55	0.83
Body mass index	26.92	4.27	26.49	3.16	26.65	3.62	0.48
Average drinks/hour[a]	2.02	0.91	2.09	0.70	2.06	0.79	0.60
Number of binges[a]	4.00	4.23	2.67	2.95	3.18	3.55	0.03
Percentage drunk[a,b]	32.27	28.71	23.59	25.39	26.94	26.97	0.04
	n	%	*n*	%	*n*	%	
Ethnicity							0.01
Hispanic/Latino	48	75	57	56	105	63	
Race							0.59
American Indian	2	3	5	5	7	4	
Asian	0	0	1	1	1	1	
Black/African	5	8	14	14	19	11	
White	43	67	55	54	98	59	
More than one race	9	14	16	15	25	15	
Unknown	5	8	11	11	16	10	

[a]In prior 28 days computed from information on the TLFB-1.
[b]Percent of time drunk when drinking.

were not significantly associated with Staff-derived values ($r = 0.119$, 95% CI = [−0.034, 0.266], $P = 0.128$).

Total binge scores ranged from 4 to 129 and from 3 to 139 for Participant- and Staff-derived scores, respectively. Agreement between Participant- and Staff-derived binge scores was evaluated using a bivariate linear regression. Agreement between these two measures was poor; the intercept significantly differed from 0 (intercept = 25.235, $P < 0.001$) and the slope significantly differed from 1 (slope = 0.299, $P < 0.001$) (Fig. 1).

Participants were categorized into Binge, Non-Binge or Unclassified drinkers based on Participant and Staff-derived binge scores using Townshend and Duka's (2005) recommendations (Table 2). Participant- and Staff-derived binge drinking classification groups were significantly associated (Fisher's exact test; $P = 0.013$), but the agreement (weighted $\kappa = 0.21$, 95% CI = [0.07, 0.36]) was generally poor and the overall concordance rate between these two classification methods was only 48%.

Staff-derived binge scores also tended to account for more variance than the Participant-derived binge scores for the categorization of drinkers in separate models predicting the number of drinks per hour (binge scores $R^2 = 0.20$; categorization $R^2 = 0.08$), number of times drunk (binge scores $R^2 = 0.21$; categorization $R^2 = 0.10$) and the percentage drunk when drinking as the outcomes (binge scores $R^2 = 0.22$; categorization $R^2 = 0.11$) (results not displayed).

Predictive validity of participant- and staff-derived drinker classification

Overall, the MANOVA yielded a non-significant effect of Participant-derived binge drinking classifications (Wilks' Lambda = 0.93, $P = 0.07$), a significant effect of Staff-derived binge drinking classifications (Wilks' Lambda = 0.83, $P < 0.001$), and a non-significant interaction between Participant- and Staff-derived binge drinking classifications (Wilks' Lambda = 0.88, $P = 0.08$) on the three AUQ items (i.e. average number of drinks per hour, number of times drunk and the percentage of times drunk when drinking). Specifically, participant-derived binge drinking classifications (i.e. Binge, Non-Binge or Unclassified) at study entry did a poor job of identifying group differences in drinking behavior observed at TLFB-2 (Fig. 2). While classification groups differed on average drinks per hour [$F(2,156) = 4.024$, $P = 0.020$], post-hoc comparisons indicated that only the Binge vs. Non-Binge contrast was significant ($P = 0.006$) (Fig. 2a). There were no differences between the binge classification groups on number of times drunk [$F(2,156) = 0.316$, $P = 0.729$; Fig. 2b] or percentage of the time drunk when drinking [$F(2,156) = 0.422$, $P = 0.657$; Fig. 2c].

In contrast, Staff-derived binge drinking classification groups showed more consistent group differences in drinking behavior observed at TLFB-2 (Fig. 2). Group differences were observed for all three drinking variables: average drinks per hour [$F(2,156) = 10.683$, $P < 0.001$; Fig. 2a]; number of times drunk [$F(2,156) = 10.809$, $P < 0.001$; Fig. 2b] and percentage of times drunk when drinking [$F(2,156) = 13.728$, $P < 0.001$; Fig. 2c], with Binge and Non-Binge drinkers being significantly different from each other on all three AUQ items (all $P < 0.001$). Unclassified drinkers differed from the Non-binge drinkers in the average number of drinks per hour ($P = 0.04$) and the percentage of the time drunk when drinking ($P = 0.02$).

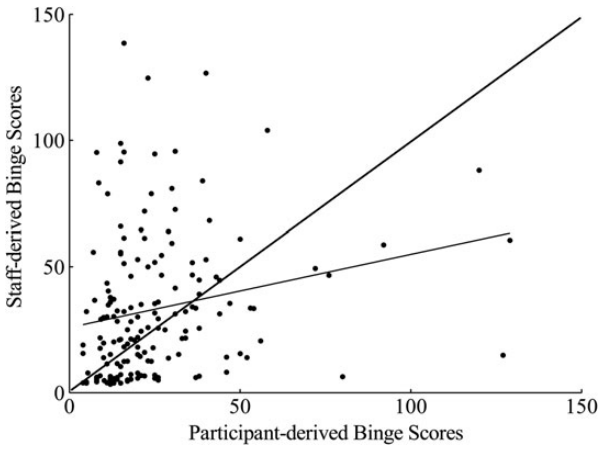

Fig. 1. Scatterplot of scores when Participant-derived binge scores predict Staff-derived binge scores. The identity line is plotted along with the regression line for reference.

Consistency of drinking variables between TLFB-1 and TLFB-2

Because the Staff-derived binge drinking classifications were determined from TLFB observations, we also examined the individual correlations of each of the three AUQ items completed by staff at study entry (TLFB-1) and subsequently measured during the study (TLFB-2) to determine the test-retest reliability of these observations. We found significant correlations between TLFB-1 and TLFB-2 on all three items: drinks per hour ($r = 0.421$, 95% CI = [0.287, 0.539], $P < 0.001$), number of times drunk ($r = 0.492$, 95% CI = [0.367, 0.599], $P < 0.001$) and percentage of time drunk when drinking ($r = 0.512$, 95% CI = [0.390, 0.616], $P < 0.001$).

ROC curves of binge scores to predict binge behaviors

To compare Participant- and Staff-derived binge scores in predicting binge behaviors, participants were rank ordered and split into quartiles based on their number of binges in TLFB-2. Separate ROC curves were calculated to compare Participant- (Fig. 3a) and Staff-derived (Fig. 3b) binge scores to predict the top (frequent bingers) and bottom (infrequent bingers) quartiles. The ROC curves revealed that Participant-derived binge scores were a better predictor of binge behavior in TLFB-2, AUC = 0.87, 95% CI [0.79, 0.95] than Staff-derived binge scores, AUC = 0.68, 95% CI [0.56, 0.81].

Table 2. Comparison of drinker classifications by Participant- and Staff-derived binge scores

		Staff-derived drinker classification			
		Non-Binge	Unclassified	Binge	Total
Participant-derived drinker classification	Non-Binge	30	8	30	68
	Unclassified	15	6	14	35
	Binge	15	4	44	63
	Total	60	18	88	166

Fisher's exact test: $P = 0.013$.

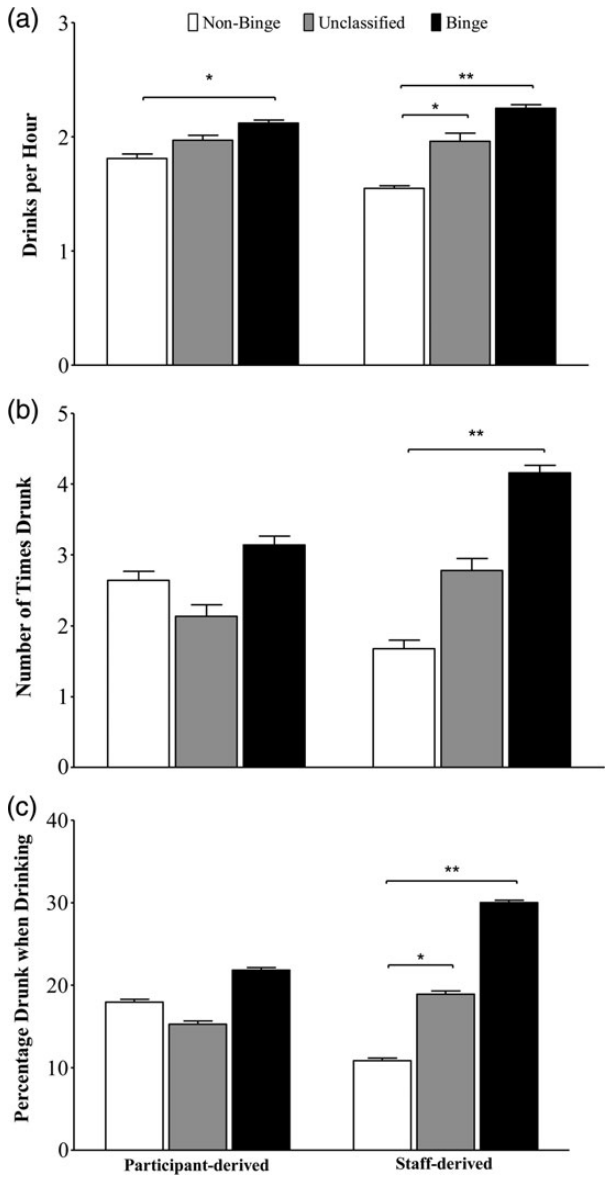

Fig. 2. Comparison of Participant-derived (left side of figures) and Staff-derived (right side of figures) drinking classifications on the three variables of interest (drinks per hour (**a**), number of times drunk (**b**) and percent of time drunk when drinking (**c**)) from TLFB-2. Data shown are means and standard errors. *indicates $P < 0.05$; **indicates $P < 0.001$.

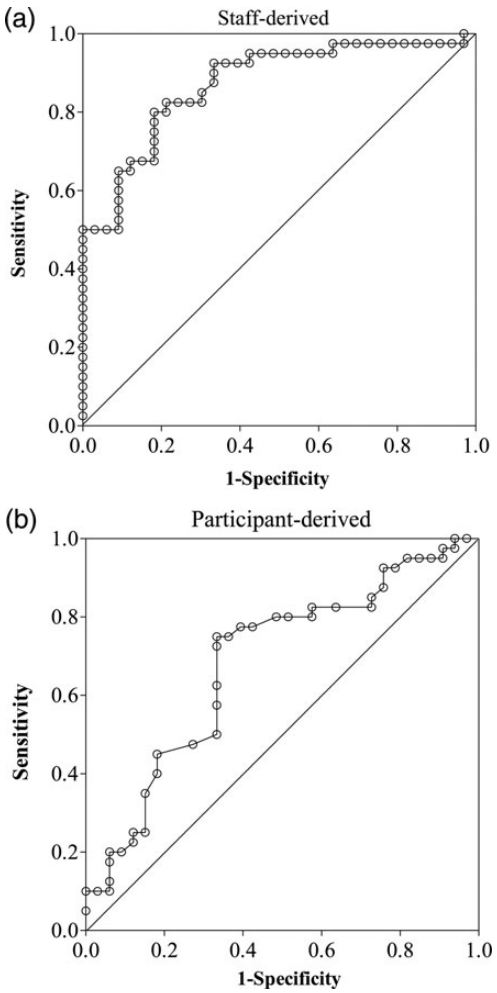

Fig. 3. ROC curves indicating the sensitivity and specificity associated with Staff- (**a**) and Participant-derived (**b**) binge scores as they predict frequent and infrequent binge drinkers. Staff-derived binge scores are better predictive of frequent and infrequent binge drinkers than Participant-derived binge scores.

DISCUSSION

The current study evaluated the value of using three questions on the AUQ (Townshend and Duka, 2005) to classify and predict binge drinking patterns in a large sample of volunteers. We employed two methods of calculating binge pattern drinking scores from the AUQ: participant-derived, where participants filled out the AUQ; and Staff-derived, where staff completed the equation using TLFB interview information. We found that: (a) Participant-derived binge scores were not convergent with Staff-derived binge scores showing overall poor classification agreement with a concordance rate of only 48%; (b) only Staff-derived binge classification had good predictive validity for the three AUQ items (drinks per hour, times getting drunk and percentage of time drunk when drinking) during study participation and (c) Staff-derived binge scores were better predictive of group classification of frequent and infrequent bingers during TLFB-2. These findings support the validity of classifying binge drinkers using the Townshend and Duka (2005) formula, but only when variables are calculated by study staff using TLFB techniques. Thus, in our study of regular pattern drinkers, self-report data on the AUQ lacked the convergent or the predictive validity suggested by others (Townshend and Duka, 2005).

Several possible reasons may explain the disparity between Participant-derived and Staff-derived binge scores and classifications. Self-report measures have typically been used to assess participant drinking because they are inexpensive and easy to administer (for review, see Del Boca and Darkes, 2003). Measures that require generalizations of typical drinking habits such as the AUQ items used to calculate Participant-derived binge scores and classifications may produce inaccurate information depending on the recall strategies used (Conrad *et al.*, 1998; Del Boca and Darkes, 2003). In comparison, calendar-based or daily estimation methods such as the TLFB interviews used to compute the Staff-derived binge scores and classifications, have well-established validity and reliability (Sobell *et al.*, 1979; Tonigan *et al.*, 1997) and are typically better assessments than those averaging general or 'typical' patterns (Sobell *et al.*, 1996; Del Boca and

Darkes, 2003). This likely explains why the computation of the binge scores in the current study derived by staff had better predictive validity than those derived using participant self reports of drinking. However, the definitions of concepts used (such as what it means to be 'drunk') to answer the AUQ items may differ between staff and participants and may contribute to differences in the predictive validity. For example, two of the three items used to derive binge scores were in relation to being drunk (the number of times drunk and the percent of times drunk when drinking). Staff, when using the TLFB to compute answers to these questions used the NIAAA (2004) definition of a binge to define a drunken episode because for a typical adult this would result in a blood alcohol level 0.08 g/dl, the legal definition of drunk in the United States. Participants, when completing the AUQ to answer these questions reported number of times drunk and percent of times drunk when drinking based on their own perception of drunk which may result in under or over reporting drunken episodes. Although we acknowledge that the estimations of the number of times and percent of times drunk when drinking in the previous six months to compute Staff-derived binge drinking scores were extrapolated from one month's TLFB data, and thus may not be a true representation of these items, research has shown that using a one-month TLFB window can reliably estimate alcohol use over longer periods of time (Vakili *et al.*, 2008).

Nonetheless, our results demonstrate that using the TLFB interviews to compute binge scores based on Townshend and Duka's (2005) recommendation have superior predictive validity compared to Participant-derived binge scores and classifications. As such, Participant-derived binge scores and classifications may be limited in their ability to assess risk or identify problematic patterns of drinking. As a result, we recommend the use of Staff-derived binge scores and classifications, using Townshend and Dukas's (2005) recommended equation. We do acknowledge, however, that using TLFB interviews to compute binge scores and classifications may not be practical for some research and clinical applications, as it can be time consuming or potentially difficult to administer correctly.

In addition, our results further show that a dichotomous classification of Binge vs. Non-Binge misses the entire group of 'Unclassified' mid-range drinkers (11% of the current sample) who show significant increases in all three binge drinking parameters relative to the non-binge group. Thus, Townshend and Duka's (2005) method of dichotomously classifying drinkers may miss individuals who engage in risky or harmful patterns of drinking, like those who were Unclassified in the current study. Alternatively, rather than using these scores dichotomously, it may be useful to use these on a continuous scale to define risk. In other words, continuous binge scores may provide an index of risk and/or drinking problems.

In summary, this study indicates that, when using the procedure developed by Townshend and Duka (2005), solely relying on Participant-derived binge scores may be problematic given their limited ability to predict actual drinking patterns. Instead, Staff-derived binge scores using TLFB techniques better relate to observed drinking because they are based on calendar-based interviews of daily patterns rather than general estimates. Lastly, categorized binge and non-binge grouping schemes may not adequately characterize the extent of problematic drinking. When scientifically justified, a continuous score or a risk index related to problematic drinking may better elucidate relationships between variables of interest.

ACKNOWLEDGEMENTS

The authors appreciate the supportive functions performed by our valued colleagues, Sharon Cates, Cameron Hunt, Krystal Shilling and Phillip Brink. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Dr Dougherty also gratefully acknowledges support from a research endowment, the William and Marguerite Wurzbach Distinguished Professorship.

FUNDING

Research reported in this publication was supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health [R01AA14988 to Donald Dougherty, Principal Investigator]. The research was also supported in part by the National Institute of Drug Abuse [T32DA031115] for postdoctoral training for Drs Lake and Karns.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

Aston ER, Shannon EE, Liguori A. (2014) Anxiety, sedation, and simulated driving in binge drinkers. *Psychol Addict Behav* **28**:359–66.

Auerbach KJ, Collins LM. (2006) A multidimensional developmental model of alcohol use during emerging adulthood. *J Stud Alcohol* **67**:917–25.

Bauer LO, Ceballos NA. (2014) Neural and genetic correlates of binge drinking among college women. *Biol Psychol* **97**:43–8.

Beseler CL, Taylor LA, Kraemer DT., *et al.* (2012) A latent class analysis of DSM-IV alcohol use disorder criteria and binge drinking in undergraduates. *Alcohol Clin Exp Res* **36**:153–61.

Borges NJ, Hansen SL. (1993) Correlation between college students' driving offenses and their risks for alcohol problems. *J Am Coll Health* **42**:79–81.

Carlson SR, Johnson SC, Jacobs PC. (2010) Disinhibited characteristics and binge drinking among university student drinkers. *Addict Beh* **35**: 242–51.

CDC. (2012) Binge drinking: Nationwide problem, local solution. http://www.cdc.gov/VitalSigns/pdf/2012–01-vitalsigns.pdf (Accessed 3 June 2013, date last accessed).

CDC. (2014) Fact sheets: Binge drinking. http://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm (Accessed on 13 November 2014, date last accessed).

Charles J, Valenti L, Miller G. (2011) Binge drinking. *Aust Fam Physician* **40**:569.

Conrad FG, Brown NR, Cashman ER. (1998) Strategies for estimating behavioural frequency in survey interviews. *Memory* **6**:339–66.

Courtney KE, Polich J. (2009) Binge drinking in young adults: data, definitions, and determinants. *Psychol Bull* **135**:142–56.

Cranford JA, Mccabe SE, Boyd CJ. (2006) A new measure of binge drinking: prevalence and correlates in a probability sample of undergraduates. *Alcohol Clin Exp Res* **30**:1896–905.

Del Boca FK, Darkes J. (2003) The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction* **98**:1–12.

Dougherty DM, Lake SL, Hill-Kapturczak N., *et al.* (2015) Using contingency management procedures to reduce at-risk drinking in heavy drinkers. *Alcohol: Clin Exp Res* **39**:743–51.

Duncan DF. (1997) Chronic drinking, binge drinking and drunk driving. *Psychol Rep* **80**:681–2.

First MB, Spitzer RL, Givvon M., *et al.* (2001) *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-Patient Edition (SCID-I/NP)*. New York: Biometrics Research, New York State Psychiatric Institute.

Gmel G, Kuntsche E, Rehm J. (2011) Risky single-occasion drinking: bingeing is not bingeing. *Addiction* **106**:1037–45.

Jackson JE, Doescher MP, Hart LG. (2006) Problem drinking: rural and urban trends in America, 1995/1997 to 2003. *Prev Med* **43**:122–4.

Mehrabian A, Russell JA. (1978) A questionnaire measure of habitual alcohol use. *Psychol Rep* **43**:803–6.

Naimi TS, Brewer RD, Mokdad A., *et al.* (2003) Binge drinking among us adults. *JAMA* **289**:70–5.

NIAAA. (2000) *Tenth Special Report to the U.S. Congress on Alcohol and Health*. Bethesda, MD: National Institutes of Health.

NIAAA. (2004) National Institute of Alcohol Abuse and Alcoholism Council approves definition of binge drinking. *NIAAA Newsl* **3**:3.

Oei TP, Morawska A. (2004) A cognitive model of binge drinking: the influence of alcohol expectancies and drinking refusal self-efficacy. *Addict Behav* **29**:159–79.

Olthuis JV, Zamboanga BL, Ham LS., *et al*. (2011) The utility of a gender-specific definition of binge drinking on the AUDIT. *J Am Coll Health* **59**:239–45.

Orchowski LM, Mastroleo NR, Borsari B. (2012) Correlates of alcohol-related regretted sex among college students. *Psychol Addict Beh* **26**: 782–90.

SAMHSA (2007) Results from the 2006 National Survey on Drug Use and Health: National Findings. In SERVICES USDOHAH (ed.). Rockville, MD: Substance Abuse and Mental Health Services Administration.

SAMHSA (2012) Binge drinking: terminology and patterns of use. Retrieved from http://captus.samsha.gov/access-resources/binge-drinking-terminology-and-patterns-use.

Shultz JM, Rice DP, Parker DL., *et al*. (1991) Quantifying the disease impact of alcohol with ARDI software. *Public Health Rep* **106**:443–50.

Sobell LC, Sobell MB. (1992) Timeline Followback: a technique for assessing self-reported ethanol consumption. In Allen J, Litten RZ (eds). *Measuring Alcohol Consumption: Psychosocial and Biological Methods*. Totowa, NJ: Humana Press, 41–72.

Sobell LC, Maisto SA, Sobell MB., *et al*. (1979) Reliability of alcohol abusers' self-reports of drinking behavior. *Behav Res Ther* **17**:157–60.

Sobell LC, Sobell MB, Leo GI., *et al*. (1988) Reliability of a timeline method: assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *Br J Addict* **83**:393–402.

Sobell LC, Brown J, Leo GI., *et al*. (1996) The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug Alcohol Depend* **42**:49–54.

Stahre M, Roeber J, Kanny D., *et al*. (2014) Contribution of excessive alcohol consumption to deaths and years of potential life lost in the United States. *Prev Chronic Dis* **11**:E109.

Tabachnik BG, Fidell LS. (2007) *Using Multivariate Statistics*, 5th edn. Boston: Allyn and Bacon.

Tonigan JS, Miller WR, Brown JM. (1997) The reliability of Form 90: an instrument for assessing alcohol treatment outcome. *J Stud Alcohol* **58**:358–64.

Townshend JM, Duka T. (2005) Binge drinking, cognitive performance and mood in a population of young social drinkers. *Alcohol Clin Exp Res* **29**:317–25.

Vakili S, Sobell LC, Sobell MB., *et al*. (2008) Using the Timeline Followback to determine time windows representative of annual alcohol consumption with problem drinkers. *Addict Behav* **33**:1123–30.

Wechsler D. (1999) *Wechsler Abbreviated Scale of Intelligence*. San Antonio, Texas: The Psychological Corporation: Harcourt Brace & Company.

Wechsler H, Dowdall GW, Davenport A., *et al*. (1995) Correlates of college student binge drinking. *Am J Public Health* **85**:921–6.